

Physics-related epistemic uncertainties in proton depth dose simulation

Maria Grazia Pia, Marcia Begalli, Anton Lechner, Lina Quintieri, and Paolo Saracco

Abstract—A set of physics models and parameters pertaining to the simulation of proton energy deposition in matter are evaluated in the energy range up to approximately 65 MeV, based on their implementations in the Geant4 toolkit. The analysis assesses several features of the models and the impact of their associated epistemic uncertainties, i.e. uncertainties due to lack of knowledge, on the simulation results. Possible systematic effects deriving from uncertainties of this kind are highlighted; their relevance in relation to the application environment and different experimental requirements are discussed, with emphasis on the simulation of radiotherapy set-ups. By documenting quantitatively the features of a wide set of simulation models and the related intrinsic uncertainties affecting the simulation results, this analysis provides guidance regarding the use of the concerned simulation tools in experimental applications; it also provides indications for further experimental measurements addressing the sources of such uncertainties.

Index Terms—Monte Carlo, simulation, Geant4, hadron therapy.

I. INTRODUCTION

THE simulation of the energy deposited by protons in matter is relevant to various experimental applications; radiotherapeutical applications exploit its peculiar pattern prior to stopping, exhibiting the characteristic “Bragg peak”, to deliver a well localized dose to the tumor area [1].

Several applications of general purpose Monte Carlo systems, like MCNP [2]–[4], GEANT 3 [5], Geant4 [6], [7], SHIELD-HIT [8], [9], FLUKA [10], [11] and PHITS [12] are documented in the literature concerning this topic, in hadron therapy as well as in other fields. A variety of physics options - theoretical models, evaluated data compilations and values of relevant physical parameters - is available in these Monte Carlo codes to model the electromagnetic and nuclear interactions of protons, and of their secondary products. While the software implementations are specific to each Monte Carlo code, the underlying physics modeling approaches and data compilations are often common to various simulation systems.

Some of the physics models and parameters used in the simulation of proton interactions with matter are affected by epistemic uncertainties [13], i.e. uncertainties due to lack of

knowledge. They may originate from various sources [14], such as incomplete understanding of fundamental physics processes, or practical inability to treat them thoroughly, non-existent or conflicting experimental data for a physical parameter or model, or the application of a physics model beyond the experimental conditions in which its validity has been demonstrated (e.g. at lower or higher energies, or with different target materials).

The role of epistemic uncertainties in the software verification and validation process has been the object of research in the context of simulation based on deterministic methods [13]; these investigations are motivated by the rigorous risk analysis required by some sensitive applications. Limited attention has been devoted so far to the role of epistemic uncertainties in Monte Carlo simulation in particle and nuclear physics, and related experimental fields. This paper addresses this topic by illustrating it in a concrete experimental use case: the simulation of proton depth dose to water for radiotherapy applications. The simulation configuration involves a realistic model of a therapeutical proton beam line and beam energies of approximately 65 MeV; this use case is representative of experimental environments for the treatment of ocular melanoma [15]–[18].

Due to their intrinsic nature, related to lack of knowledge, epistemic uncertainties are difficult to quantify [19]. Although the characterization of epistemic uncertainty contributions is needed for many of the issues that feed the reliability model of complex systems, there is no generally accepted method of measuring epistemic uncertainties and their contribution to reliability estimations. A variety of mathematical formalisms [20] has been developed for this purpose; nevertheless, some of the techniques adopted in the context of deterministic simulations, like interval analysis and applications of Dempster-Shafer theory of evidence [21], are not always directly applicable in identical form to the treatment of epistemic uncertainties in Monte Carlo simulations.

Sensitivity analysis [19] is a tool for exploring how different uncertainties, including epistemic ones, influence the model output [22]. This approach is adopted in the study described here: the paper identifies a set of epistemic uncertainties in physics modeling pertinent to the problem domain, documents their impact on the simulation results, and discusses their potentiality to produce systematic effects in relation to the characteristics of the application environment.

Typically, in statistical analyses epistemic uncertainty is represented as a set of discrete possible or plausible choices (e.g. model choices) [23]. The environment for this kind of study has been realized in the context of a Geant4-based

Manuscript received March 30, 2010.

M. G. Pia and P. Saracco are with INFN Sezione di Genova, Via Dodecaneso 33, I-16146 Genova, Italy (phone: +39 010 3536328, fax: +39 010 313358, e-mail: MariaGrazia.Pia@ge.infn.it, Paolo.Saracco@ge.infn.it).

M. Begalli is with State University of Rio de Janeiro, Rio de Janeiro, Brazil (email: Marcia.Begalli@cern.ch).

A. Lechner is with the Atomic Institute of the Austrian Universities, Vienna University of Technology, Vienna, Austria and CERN, Geneva, Switzerland (email: Anton.Lechner@cern.ch).

L. Quintieri is with INFN Laboratori Nazionali di Frascati, Frascati, Italy (e-mail: Lina.Quintieri@lnf.infn.it).

application; the characteristic of Geant4 as a toolkit, encompassing a wide variety of physics models, along with the feature of polymorphism characterizing the object oriented programming paradigm, allow the configuration of the simulation with a large number of different physics options in the same software environment. This versatility makes Geant4 a convenient playground to evaluate the effects of a number of physics alternatives on the experimental use case under study. The sensitivity analysis documented in this paper, which examines the response of the system to a wide set of modeling approaches, plays a conceptually similar role to the interval analysis method applied in deterministic simulation, where parameters subject to epistemic uncertainties are varied within bounds. As mentioned in the previous paragraph, sensitivity analysis as applied to the context described in the following sections contributes to identify and quantify possible systematic effects in the simulation; it cannot infer anything about the validity of any of the physics models, for which experimental data would be needed.

The physics modeling options available in Geant4 for the use case under study are broadly representative of the body of knowledge in the problem domain; the analysis described in this paper, albeit performed in the context of a specific Monte Carlo system, provides elements of interest for other simulation environments as well.

Epistemic uncertainties are in principle reducible [13]; the analysis documented in this paper identifies areas where experimental measurements could reduce them, and highlights their requirements of accuracy and other features to improve the knowledge embedded in the current simulation models.

Preliminary assessments relevant to this study were reported in [24].

II. GEANT4-BASED SIMULATION OF PROTON DEPTH-DOSE PROFILES

The Geant4 toolkit includes various physics modelling options relevant to the application domain considered in this paper; they concern stopping powers, multiple scattering, cross sections and final state models of elastic and inelastic nuclear interactions of the primary protons, as well as a variety of electromagnetic and hadronic models for the secondary particles resulting from proton interactions with matter.

Several Geant4-based simulations of proton therapy set-ups, like [25]–[35], have documented satisfactory agreement with experimental depth dose measurements in various configurations of beam lines and detectors. Some open issues remain, which are generally shared by simulations based on other Monte Carlo systems as well, due to common physics modeling approaches in the codes and similar practices in the experimental domain.

There is evidence in the literature of different features produced by Geant4 physics models in the energy range below 100 MeV [36]–[40]; the discrimination of their accuracy is made difficult by the still incomplete software validation, which is often hindered by the limited availability of experimental data, or their controversial characteristics. The same problem affects other Monte Carlo codes as well. Furthermore,

some shortcomings are present in the comparisons of proton therapy simulations with experimental data.

The physics configuration used in the simulation and the Monte Carlo kernel version are either undocumented, or incompletely documented, in a number of references; therefore it is not always possible to relate the results documented in the literature to the physics options and software implementations which produced them, thus hindering the reproducibility of the results. The comparison between simulated and experimental data is limited to qualitative appraisal in most cases; the lack of statistical analysis in many articles prevents the reader from appraising the significance of the compatibility between simulation and experimental data, and the relative merits of the simulation models in comparison to measurements taken in different experimental configurations and characterized by different uncertainties.

It is a feature of proton therapy simulations that the beam parameters, namely the beam energy and energy spread, are not known in typical experimental set-ups of this domain with sufficient accuracy to base the simulation on their nominal values [27]. The precise knowledge of the beam energy is not critical to clinical applications; in that context what is of interest is the proton range, which can be measured very accurately. In common practice, the beam parameters input to the simulation are adjusted so that the simulation results best fit the measured depth dose profile [25], [27], [30], [33], [41], [42]; this procedure affects the significance of further comparisons between simulated and experimental data. Experimental techniques have been developed to measure the energy of a proton beam from radiotherapy accelerators with greater precision [43], but they are not commonly exploited in connection with simulation validation studies.

The comparison of simulated and experimental Bragg peak profiles is also sensitive to the normalization procedures, which are often applied in experimental practice to relate simulated and measured data; this topic is discussed in detail in [44].

These shortcomings do not severely affect current clinical practice, where Monte Carlo simulation plays a role of verification and optimization. More demanding requirements about the predictive capabilities of the simulation may derive from new perspectives, such as the use of Monte Carlo simulation in treatment planning, which is the object of ongoing research [45]–[48], or other disciplines, like radiation protection.

Epistemic uncertainties undermine the predictive role of simulation software; by identifying and quantifying them, the analysis described in the following sections is propaedeutic to further experimental and theoretical efforts to reduce them, or at least to control their effects.

III. OVERVIEW OF RELEVANT PHYSICS MODELS

A brief overview of the Geant4 physics models relevant to the use case addressed in this paper is given below to facilitate the understanding of the results presented in the following sections. In the context of Geant4, particle interactions with matter are represented by processes, which can be implemented through different models [6]. Similarities

with the modeling approaches in other Monte Carlo codes are discussed; they show that the body of knowledge, as well as knowledge gaps, are largely shared across a variety of simulation systems. The information summarized in this section is necessarily succinct; further details can be found in the cited references.

A. Electromagnetic interactions

Geant4 electromagnetic package [49] includes two packages relevant to the experimental use case considered: *Standard* and *Low Energy*.

The proton ionisation process in the *Low Energy* electromagnetic package [50], [51] is articulated through models [52] based on the free electron gas model [53] at lower energy (<1 keV), on parameterisations at intermediate energy, and on the Bethe-Bloch equation at higher energy. Alternative parameterisation models, identified in the following as *ICRU49*, *Ziegler77*, *Ziegler85* and *Ziegler2000* implement electronic and nuclear stopping powers based on [54]–[57]; their different energy ranges of applicability are documented in the respective references. The configuration of the hadron ionisation process is identified in the following sections through the selected parameterisation model. The process also deals with the emission of δ -rays.

The Geant4 *Low Energy* electromagnetic package includes processes [58], [59] to handle the secondary particles resulting from proton interactions: electrons, photons and ions. Models for electrons and photons are based on data libraries (EEDL [60] and EPDL97 [61]) and on analytical formulations originally developed for the Penelope [62] Monte Carlo code; both options account for the atomic relaxation [63] following the primary processes. Models based on the parameterisations in [55], [56], and on [64] are available for ions.

The main features of the Geant4 *Standard* electromagnetic package are documented in [65]. More recently a model [68] for proton ionisation based on [54] was implemented in this package; this physical approach is the same as the one already present in the *Low Energy* package. The handling of energy loss fluctuations is implemented in this package; it is based on the model adopted in GEANT 3 [66] and was updated in Geant4 8.3 [67].

An assessment of Geant4 electromagnetic processes against the NIST (United States National Institute of Standards) reference data can be found in [36]. The validation of Geant4 low energy electromagnetic processes against precision measurements of electron energy deposition is documented in [69].

The *Standard* package includes implementations [70], [71] of the multiple scattering process. In the early Geant4 versions a generic process (*G4MultipleScattering*), based on the Lewis [72] theory, was applicable to any charged particles; it has been complemented by a process specialized for hadrons (*G4hMultipleScattering*) in Geant4 8.2, and one specific to electrons (*G4eMultipleScattering*) in Geant4 9.3. The specialized multiple scattering processes are intended to replace the generic one in future Geant4 releases [73]. These processes can be configured with various models (e.g. *G4UrbanMscModel90*, *G4UrbanMscModel92* and

G4UrbanMscModel93), which involve some empirical parameters [74]. Early implementations of Geant4 multiple scattering were validated against experimental muon data at low [75], [76] and high [77] energy; [77] showed better accuracy of Geant4 multiple scattering model (as implemented in Geant4 1.0) with respect to the GEANT 3 model based on Molière theory. Recent studies, like [69], [78], have highlighted issues in the evolution of energy deposition patterns involving electron-photon interactions simulated with different Geant4 versions; the observed effects have been ascribed to modifications to Geant4 multiple scattering algorithm. The literature concerning recent evolutions of Geant4 multiple scattering is focused on issues related to the dependence of the simulation on transport step size and parameters of the algorithm [67], [68], [74]. To the best of the authors' knowledge, the validation of Geant4 proton multiple scattering is not documented in literature yet for the energy range relevant to this study.

The Geant4 *Standard* and *Low Energy* electromagnetic packages are concerned by the same epistemic uncertainties affecting electromagnetic physics modeling relevant to the use case studied in this paper, which are analyzed in sections V-B and V-C. Differences in the outcome of simulations using the two packages may derive from features specific to each package, which are not associated with epistemic uncertainties in the physics domain pertinent to the use case examined in this paper; they could be due to numerical features, like the number of bins in the look-up tables used in the simulation and their interpolation, or algorithms and modelling choices specific to each package. A complete documentation and analysis of different features of Geant4 *Standard* and *Low Energy* electromagnetic packages is outside the scope of this paper.

Other Monte Carlo codes used for hadron therapy simulation adopt similar approaches for stopping power calculation at high and low energies. At intermediate energies, stopping powers based on ICRU 49 Report are implemented in SHIELD-HIT; an improvement to include them in FLUKA is documented in [79], but it does not appear to be released yet in the current version of FLUKA (FLUKA-2008) at the time of writing this paper. PHITS handles proton ionization according to the SPAR code [80], while MCNP uses an energy weighted average between the high and low energy calculations [81], which adopt the same methodology as in SPAR. PHITS [82], SHIELD-HIT [42] and MCNPX do not model δ -ray emission.

GEANT development was frozen with the 3.21 version in 1994; the code is no longer supported, but it is still used for hadron therapy developments [83]. GEANT 3 simulated proton energy loss based on the Bethe-Bloch equation and dealt with δ -ray production from ionization.

MCNP multiple Coulomb scattering treats soft and hard interactions separately [84]: soft collisions are described using a continuous scattering approximation; a small number of hard collisions are simulated directly. Multiple scattering is based on Molière theory in FLUKA and PHITS [85]; details of the algorithm in FLUKA are described in [86]. SHIELD-HIT simulates multiple scattering on the basis of a gaussian model [42], which gives the correlated value of the angular deviation and lateral displacement of the scattered particle. GEANT 3

provided two options for multiple scattering simulation, respectively based on a Gaussian approximation and on Molière theory [5].

B. Hadronic interactions

The Geant4 hadronic package addresses the complexity of nuclear interactions through a software framework [87]. The baseline design can accommodate multiple implementations of cross sections and final state models associated with a process, which are either complementary in their energy range coverage or alternative in their modelling approach; processes and models are meant to be handled polymorphically through their respective base classes.

1) *Elastic scattering*: Geant4 includes various elastic scattering processes: *G4HadronElasticProcess*, *G4UHadronElasticProcess*, *G4QElastic* and *G4WHadronElasticProcess*; the latter was released in Geant4 9.3 with the purpose [73] of allowing models for elastic scattering to be treated in a similar way to inelastic models. A *G4DiffuseElastic* [88] process was also released in Geant4 9.3; the energy range of its applicability is not explicitly specified in [88], but, since this process appears applied at energies of 1 GeV and above in the associated reference, one is led to assume that it is pertinent to higher energies than those relevant to the use case studied in this paper. This inference manifests an epistemic uncertainty in the applicability domain of this process.

The *G4HadronElasticProcess* class of the hadronic *processes* package handles cross section and final state calculation according to the software design of [87].

It can be configured with the *G4HadronElasticDataSet* class, derived from *G4VCrossSectionDataSet* and included in the Geant4 hadronic *cross_sections* package, which implements total elastic scattering cross sections derived from GHEISHA [89].

The scattering can be configured through several models, such as *G4LElastic*, *G4ElasticCascadeInterface* and *G4HadronElastic*. *G4LElastic*, included in the hadronic *models* package, is based on GHEISHA algorithms reengineered in Geant4; it is not meant to conserve energy and momentum on an event-by-event basis, but only on average. *G4ElasticCascadeInterface*, identified in the following as “Bertini elastic”, is included in the *cascade* package of the *cascade* package of hadronic *models*; it derives from *G4IntraNuclearTransportModel* and implements a model based on the INUCL [90] code. The *G4CascadeElasticInterface* class in the same package activates both elastic and inelastic interactions. *G4HadronElastic* [91], included in the *coherent_elastic* package of hadronic *models*, combines elements originally developed for CHIPS (Chiral Invariant Phase Space) [92], [93] with other modelling approaches; it aggregates a *G4VQCrossSection* object belonging to CHIPS and a *G4ElasticHadrNucleusHE* model.

The *G4QElastic* process, known as “CHIPS elastic”, delegates the calculation of cross sections to the *G4QElasticCrossSection* class, derived from *G4VQCrossSection*, and implements its own scattering

algorithm. All the related classes are included in the *chiral_inv_phase_space* package of hadronic *models*.

The *G4UHadronElasticProcess* process [91], included in the *coherent_elastic* package of hadronic *models*, is meant to be configured with the dedicated *G4HadronElastic* model; this configuration is referred to in the following as “U-elastic”.

The limited documentation in the literature of Geant4 elastic scattering models and other codes does not facilitate the appreciation of their characteristics, nor the identification of the experimental data with respect to which some of the simulation models, especially those implementing parameterisations, may have been calibrated. Although improvements to Geant4 elastic scattering modeling are mentioned in [40], hardly any validation of the available models is documented in the literature in the energy range relevant to the use case under study. The use of these models in the simulation is a source of epistemic uncertainty, due to the lack of knowledge of their accuracy in the energy range pertinent to this study.

2) *Non-elastic interactions*: Inelastic hadron-nucleus scattering is handled in Geant4 through processes specific to each particle type. Processes for protons, neutrons, deuteron, triton and α particles are relevant to this study.

Total inelastic cross sections derived from GHEISHA [89] are available in Geant4 through *G4HadronInelasticDataSet* for all hadrons, α particles, deuteron and triton; alternative implementations based on [94], [95] are available for some energy ranges and target materials. Cross sections describing neutron-nucleus scattering with higher precision below 20 MeV are available in *G4NeutronHPInelasticData*. Specialized cross sections, based on [96]- [101], are available for ions.

Parameterised and theory-driven [102] models of nuclear inelastic scattering are available in Geant4 for protons and other particles, concerning the energy range pertinent to this study.

Geant4 Low Energy Parameterised models (LEP), originating from GHEISHA [89], handle protons, neutrons, pions, α particles, deuterons and tritons.

The CHIPS *G4QInelastic* inelastic process [92], [93] implemented in Geant4 is applicable to hadron inelastic scattering in the energy range pertinent to this study.

Various options of theory-driven models describe the phases of intranuclear transport, preequilibrium and nuclear deexcitation in Geant4. Other Monte Carlo codes used in proton therapy applications use a similar multi-stage approach to simulate proton inelastic interactions. The primary proton energy in this study lies at the edge of what is commonly considered as the range of transition between intranuclear cascade and preequilibrium models.

Geant4 includes three cascade models, each one with further associated models describing the lower energy phases: they are known as Binary [103], Bertini [104], [105] and Liège [106] cascade models.

The Binary cascade model [103] adopts a hybrid approach between a classical cascade code and a quantum molecular dynamics model. It handles the preequilibrium phase through the Precompound model (*G4PreCompoundModel*) [107], whose implementation is based on Griffin’s exciton model [108],

[109]; this model can be activated in a simulation application either through the Binary cascade model or as an independent model. The nuclear deexcitation associated with the Precompound model can be configured with various options: they include an evaporation model based on Weisskopf-Ewing's [110], [111] theory, exploiting Dostrovsky's computational approach [112], the Generalized Evaporation Model (GEM) [113] (also used by PHITS), and the optional activation of the Fermi break-up [107].

A similar approach is adopted in SHIELD-HIT, whose default model considers a fast cascade stage, which brings the interaction between the projectile and target to a sequence of binary collisions [114]; this stage is followed by preequilibrium [115] and deexcitation of residual nuclei, with Fermi break up of light nuclei and evaporation.

The Geant4 Bertini Cascade implementation [104], [105] is a reengineering of the INUCL code [90], which is based on Bertini's approach [116] to intranuclear transport; it handles the preequilibrium phase based on Griffin's exciton model and the evaporation phase based on Weisskopf's and Dostrovsky's approach. The preequilibrium part of INUCL is based on the Cascade Exciton Model (CEM) [117], which is one of the options of MCNPX for proton transport and is also implemented in SHIELD (as well as in other codes) [118].

A cascade model based on Bertini's scheme, derived from the HETC [119] implementation in LAHET [120], is available in MCNPX to handle protons, besides the ISABEL [121], [122] and CEM options. The HETC model was also interfaced to GEANT 3 through CALOR [123].

The INCL Intranuclear Cascade [106] model, better known as Liège Cascade, has been reengineered in Geant4 together with the associated ABLA evaporation model [124]; it was first released in Geant4 version 9.1 [125] and further improved in Geant4 9.3. INCL is included [126] in the LAHET [120] code system used by MCNPX. Although, according to [125], INCL is meant for energies above 200 MeV, satisfactory applications at energies of a few tens MeV are reported in the literature [127]–[129].

The simulation of low energy proton interactions in PHITS is based [130] on the MCNP4C and NMTC [131] codes; the latter incorporates Bertini's cascade model [116] for nucleon and meson transport.

FLUKA handles inelastic scattering through PEANUT [132] in the energy range relevant to the use case under study; it involves a sequence of intranuclear cascade followed by preequilibrium and deexcitation. The preequilibrium model is based on the formalism developed by Blann [133] with some modifications [134]; the evaporation model is based on Weisskopf-Ewing's approach [135]; Fermi break-up is modeled for light nuclei [135].

Hadronic interactions were not handled by GEANT 3; their treatment was delegated to external packages (GHEISHA, CALOR and an early version of FLUKA) interfaced to it.

Limited documentation regarding the validation of Geant4 inelastic scattering models relevant to the use case of this study is available in the literature. Some comparisons with experimental data are reported in [38], [39], [40], [91], [103], [136], [137]: although most of the results shown in these

references are not directly related to the use case investigated in this paper, they demonstrate the ongoing validation efforts in this domain.

C. Epistemic uncertainties in the simulation models

Epistemic uncertainties in the physics simulation models arise from various sources.

In some cases, the lack of knowledge concerns the value of a physical parameter: this is the case, for instance, for the mean water ionization potential, for which various values, originating from experimental measurements or theoretical calculations, are documented in the literature.

Other sources of uncertainty are associated with values, used in the simulation, deriving from parameterisations, or fits to experimental data (which may be inspired by theoretical motivations): in the present study this concerns, for instance, the cross sections of nuclear elastic and inelastic interactions, and proton stopping powers. In these cases the uncertainties derive from the measurements themselves, the criteria by which the data are selected for the fit, and from the parameterisation or fitting process.

Some models may embed parameters or, more generally, features, which are adjusted in the software implementation according to empirical procedures: from calibration with respect to experimental data to educated guesses, in the absence of pertinent measurements. This is the case, for instance, for the Geant4 multiple scattering model and for some hadronic interaction models. In this respect, and also for models based on parameterisations or fits to experimental data, an important issue is the distinction between the calibration and validation of simulation models; for the reader's convenience, these concepts pertaining to simulation epistemology are briefly recalled here. Calibration is the process of improving the agreement of a code calculation with respect to a chosen set of benchmarks through the adjustment of parameters implemented in the code [19]; in the Monte Carlo simulation jargon this process is also known as "tuning". Software validation is defined in the IEEE Standard for Software Verification and Validation [138]. This generic definition is adapted to specific application domains with some slight variants; regarding simulation, validation is usually intended as the process of determining the degree to which a model is an accurate representation of the real world from the perspective of its intended uses, or of confirming that the predictions of a code adequately represent measured physical phenomena [19], [139].

The limited documentation in the literature of the calibration of the physics models implemented in Monte Carlo codes does not facilitate the understanding whether some of the comparisons with experimental data reported in the literature document calibration results and their experimental benchmarks, or model validation.

The use of a simulation code for predictive purposes outside the scope of its validation necessitates extrapolation beyond the understanding gained strictly from experimental validation data. This type of uncertainty in our inference is primarily epistemic.

Regarding the energy range of nuclear interactions relevant to the use case considered in this paper, a long debate has

TABLE I
PROTON PHYSICS MODELLING OPTIONS IN THE SIMULATION

| Physics domain | Option | Process class | Model class |
|-----------------------------------|--|--|--|
| Proton stopping powers | ICRU49 Ziegler77 Ziegler85 Ziegler2000 | G4hLowEnergyIonisation G4hLowEnergyIonisation G4hLowEnergyIonisation G4hLowEnergyIonisation | G4hICRU49p G4hZiegler1977p G4hZiegler1985p G4hSRIM2000p |
| Multiple scattering | Generic multiple scattering Hadron multiple scattering | G4MultipleScattering G4hMultipleScattering | G4UrbanMscModel192 G4UrbanMscModel190 |
| Hadronic elastic scattering | LEP U-elastic Bertini-elastic CHIPS-elastic | G4HadronElasticProcess G4UHadronElasticProcess G4HadronElasticProcess G4QElastic | G4LElastic G4HadronElastic G4ElasticCascadeInterface |
| Hadronic inelastic cross sections | Default Wellisch [94] | G4ProtonInelasticProcess G4ProtonInelasticProcess | G4HadronInelasticDataSet G4ProtonInelasticCrossSection |
| Hadronic inelastic scattering | LEP Precompound Precompound-GEM Precompound-Fermi break-up Binary cascade Bertini cascade Liège CHIPS-inelastic | G4ProtonInelasticProcess G4ProtonInelasticProcess G4ProtonInelasticProcess G4ProtonInelasticProcess G4ProtonInelasticProcess G4ProtonInelasticProcess G4ProtonInelasticProcess G4QInelastic | G4LEProtonInelastic G4PreCompoundModel G4PreCompoundModel, G4ExcitationHandler G4PreCompoundModel, G4ExcitationHandler G4BinaryCascade G4CascadeInterface G4InclAblaCascadeInterface |

been going on for decades in the literature about different theoretical preequilibrium models, respectively based on the so-called “exciton” and “hybrid” approaches [140]. These discussions involve subtle theoretical arguments; however, it has been acknowledged that, whichever theoretical approach is chosen to model equilibrium emission, the effects of overly simple or untested assumptions can be compensated by means of other uncorrelated phenomenological parameterisation of the model. This ongoing theoretical debate is not expected to be relevant to the use case addressed in this study, since the differences of the various theoretical approaches are expected to affect mainly exclusive channels [141], with negligible effects on the resulting deposited energy, which is the object of this paper. The sources of epistemic uncertainties in the simulation reside in the phenomenological features of the nuclear model implementations, rather than in the choice of theoretical approach to preequilibrium modeling.

The analysis described in the following sections identifies sources of epistemic uncertainties in the physics domain of the simulation and evaluates systematic effects on the simulation outcome associated with them.

IV. SOFTWARE CONFIGURATION

A. Simulation

The simulation application includes components responsible for the configuration of the geometry and materials of the experimental set-up, the generation of the proton beam, the selection of the physics features to be used in the particle transport, the collection of relevant observables in dedicated objects, and the control of the user’s interaction with Geant4 kernel at various stages of the execution.

The geometry configuration encompasses a realistic model of a proton therapy beam line and a volume, placed at the exit of it, where the energy deposited by the proton beam is scored. The beam line model exploits the code of the geometry and material composition of a real-life proton therapy facility [18], which is publicly released in the Geant4 *hadrontherapy*

[30] example; the implementation of the beam line geometry as in Geant4 8.1 was used for all the simulation productions. The scoring volume consists of 4 cm cube, filled with water; its size is adequate to contain the formation of the Bragg peak of energy loss produced by the proton beam. This volume is defined “sensitive” [6] in Geant4 terms. A readout geometry, longitudinally segmented in 200 μm thick slices, is superimposed to the mass geometry of the sensitive volume; the longitudinal segmentation determines the resolution of the simulation in the location of the Bragg peak, and mimics the resolution of typical measurements with ionization chambers in experimental practice. The energy deposit profile is scored through Geant4 hits objects. The figures of longitudinal energy deposition profiles included in the following sections show the energy deposited in each slice of the readout geometry by primary protons and secondary particles, integrated over all the events in the simulation run.

Primary protons are generated according to a Gaussian distribution with 63.95 MeV mean energy and 300 keV standard deviation, unless differently specified in the following sections.

A variety of physics options is configured by means of a software design exploiting Geant4 *G4VModularPhysicsList* and *G4VPhysicsConstructor* as base classes. A class derived from Geant4 kernel’s *G4VModularPhysicsList* is responsible for driving the physics selection in the simulation application: it assembles a physics configuration by means of subclasses of *G4VPhysicsConstructor*, each one responsible for instantiating the physics processes and models pertinent to a particle type (or a group of particles, like ions) and an interaction type (electromagnetic, hadronic elastic or hadronic inelastic).

The proton physics options and corresponding Geant4 classes evaluated in this study are summarized in Table I.

For convenience, a subset of *G4VPhysicsConstructor* subclasses, corresponding to the selection in Table II, has been defined as a reference configuration for this study.

In all the productions the interactions of secondary particles are simulated as in Table II, unless differently specified in the following sections.

TABLE II
REFERENCE PHYSICS CONFIGURATION IN THE SIMULATION

| Particles | Physics selection |
|--------------------------------|--|
| Electrons Photons | Low energy electromagnetic package, library-based models |
| Positrons | Standard electromagnetic package |
| Protons | <i>G4hLowEnergyIonization</i> , <i>ICRU49</i> parameterisations <i>G4UHadronElastic</i> process, <i>G4HadronElastic</i> model <i>G4ProtonInelasticProcess</i> , Precompound model Inelastic cross sections as in [94] |
| Neutrons | <i>G4UHadronElastic</i> process, <i>G4HadronElastic</i> model <i>G4NeutronInelasticProcess</i> , Precompound model <i>G4HadronCaptureProcess</i> , LEP model <i>G4HadronFissionProcess</i> , LEP model Inelastic cross sections as in [94] |
| Deuteron Triton α | <i>G4hLowEnergyIonization</i> , <i>ICRU49</i> parameterisations (scaled) <i>G4UHadronElastic</i> process, <i>G4HadronElastic</i> model Inelastic process specific to each particle, LEP models Tripathi and Shen cross sections |
| Charged | <i>G4MultipleScattering</i> process |

Options regarding the water ionization potential can be introduced in the simulation by setting the value of this parameter through the public interface of the *G4Material* class.

The secondary particle production threshold is set at 1 μm (in range) in the water volume and 10 μm in the passive components of the experimental set-up. The results presented in the following sections are generated with 10 μm maximum step size; the step limit is set approximately an order of magnitude smaller than the thickness of the slices in the readout geometry to ensure adequate precision in scoring the longitudinal energy deposition profile.

B. Simulation production

Simulated data were produced by instantiating various combinations of physics constructor objects in the simulation application corresponding to the options listed in Table I.

The simulation production concerned a set of representative physics configurations, each differing from the reference one of Table II by one modeling feature only; this strategy allowed the unambiguous attribution of the differences observed in each simulation to the physics feature specifically under investigation. One million events were generated for each physics configuration.

Data samples corresponding to all the options listed in Table I were produced with Geant4 9.3, the latest version available at the time of writing this paper. Data samples corresponding to a subset of physics configurations were also produced with three other Geant4 versions: 8.1p02, 9.1 and 9.2p03. Versions 8.1p02 and 9.1 were involved in the validation of electron energy deposition in [69]; Geant4 8.1 was previously used for the assesment of some Geant4 physics models in [35] (subject to the same beam settings adjustments as in [27]), while Geant4 9.1 is still widely used in production mode by various experiments. Geant4 9.2p03 is an updated version of Geant4 9.2 including corrections; it was released two months later than Geant4 9.3.

The simulations were run on Intel Core2 Duo Processors E6420, equipped with 2.13 GHz CPU and 4 GB RAM, Scientific Linux 4 operating system and gcc 3.4.6. compiler.

C. Data analysis

Data analysis objects were handled in the simulation application code through AIDA [142] abstract interfaces; the iAIDA [143] implementation was used.

The results of the various simulation configurations were compared to the outcome of the reference one. No normalization was performed on the Bragg peak profiles to be compared (unless explicitly stated in the following sections): therefore the distributions, which originate from the same number of primary protons in all the simulation configurations, allow the appreciation not only of differences in shape and peak location, but also of absolute effects, like the proton acceptance in the sensitive volume and the total energy deposited in it.

Within a given Geant4 version, any observed difference in the results is to be ascribed to intrinsic effects of the physics models activated in the simulation. To the best of the authors' knowledge, the Geant4 transport kernel did not modify its behavior through the versions considered in this study; differences observed in comparisons across different Geant4 versions reflect the evolution of the physics model implementations.

The differences associated with the various simulation configurations were quantitatively estimated by means of statistical methods; the comparisons exploited non-parametric, unbinned goodness-of-fit tests available in the Statistical Toolkit [144], [145]. Different algorithms were applied to each test case to evaluate possible systematic effects on the results due to particular features of the tests: the Anderson-Darling [146], [147], Cramer-von Mises [148]–[150] and Kolmogorov-Smirnov [151], [152] tests. The null hypothesis in the test process assumed the equivalence between the distributions subject to comparison. The critical value for the rejection of the null hypothesis was set at 0.05, unless differently specified; in the following sections the expression “95% confidence level” is used to indicate 0.05 significance of the test.

Goodness-of-fit tests compared the whole longitudinal distributions of energy deposition, as well as the distributions corresponding to the left and right branches of the longitudinal profiles, i.e. at penetration depths respectively up to and beyond the peak position, to ascertain the compatibility of the data samples in detail.

The comparison results mentioned in the following sections concern the left branch of the energy deposition profiles, unless differently specified. Due to the mathematical features of the test statistics and empirical distribution functions, the left branch of the profiles provides the most sensitive test case to highlight differences in the distributions subject to comparison.

V. RESULTS

The following sections summarize the main results of the analysis of the various Geant4 physics modelling options; they concern the longitudinal energy deposition profile.

The lateral energy deposition pattern is also of interest to proton therapy, and to other experimental applications as well; related major sources of epistemic uncertainties in the simulation are the models of multiple scattering and nuclear elastic scattering, and secondary particle production from nuclear

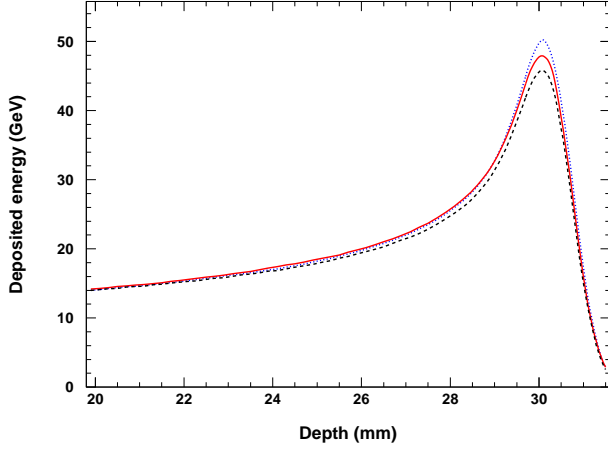


Fig. 1. Bragg peak profile resulting from electromagnetic interactions only (blue dotted curve), electromagnetic interactions and hadronic elastic scattering (solid red curve), electromagnetic, hadronic elastic and inelastic scattering (dashed black curve). The profiles were produced with the Geant4 9.3 physics configuration listed in Table II and subsets of it. The Bragg peak is from one million primary protons, generated with $\langle E \rangle = 63.95$ MeV mean energy and $\sigma = 300$ keV standard deviation incident on water; the plot shows the deposited energy in each slice of the longitudinally segmented readout geometry, integrated over all the generated events.

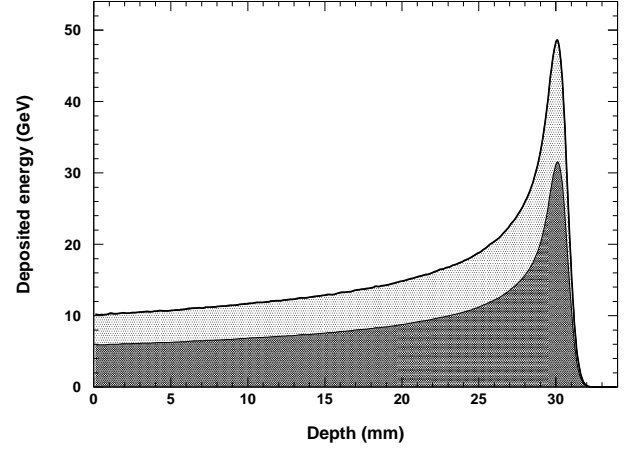


Fig. 2. Longitudinal energy deposition due to protons (dark shaded area) and electrons (light shaded area); the thick line represents the total energy deposited by all particle species. The profiles were produced with the Geant4 9.3 physics configuration listed in Table II; the electron production threshold was equivalent to 250 eV in water. The Bragg peak is from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water; the plot shows the deposited energy in each slice of the longitudinally segmented readout geometry, integrated over all the generated events.

interactions. The data samples produced for the study of the longitudinal energy deposition profile are of inadequate size to draw any statistically significant conclusions concerning the effects of epistemic uncertainties on the lateral energy distribution patterns; their quantification at comparable significance level would require simulation samples at least two orders of magnitude larger for the analysis of the lateral distribution of energy deposition than for the longitudinal one. Such a large scale simulation production in a realistic experimental use case was beyond the practical reach of the limited computational resources available to the authors; it should be pursued once adequate computing means are available.

Unless differently specified in the text, the results were produced with Geant4 version 9.3.

A. Shaping the longitudinal energy deposition

Electromagnetic, hadronic elastic and inelastic interactions contribute to a different extent to shaping the energy deposition as a function of penetration depth; hadronic interactions are known [28], [29] to be relevant to the simulation of the proton Bragg peak. The relative contribution was evaluated by activating partial and full components of the reference physics configuration in the simulation: electromagnetic interactions only, electromagnetic interactions and hadronic elastic scattering, and the full set, including hadronic inelastic processes, as in Table II. Fig. 1 shows the longitudinal energy deposition resulting from the various contributions: the peak depth and the overall pattern of energy deposit are dominated by the electromagnetic component, while elastic and inelastic hadronic interactions contribute to modify its shape. Other options of Geant4 electromagnetic and hadronic models result in similar apportioning among the various physics contributions.

The energy deposited in the sensitive volume derives from protons, electrons and other particles; the contribution from the latter amounts to less than 1%.

The relevance of electrons' contribution to the deposited energy is related to the electron production threshold set in the simulation application; the results described in this paper were produced with a threshold equivalent to 250 eV in the sensitive water volume, that is the lowest energy recommended for use with Geant4 library-based electron and photon processes, and corresponds to the setting in the validation study of [69]. The resulting contribution of secondary electrons to the longitudinal energy deposit profile is illustrated in Fig. 2.

The accuracy of the secondary electron simulation contributes to the overall accuracy of the energy deposition deriving from primary protons; epistemic uncertainties in the electron simulation models may affect the results. For Geant4, the validation of the longitudinal energy deposition of electrons in an energy range relevant to this study is documented in [69].

Low-energy electrons are of great importance in therapeutic particle beams, since they are very powerful in causing lethal damage to the cells [153]; the accuracy of δ -ray simulation is especially important for studies of the biological effects of proton irradiation.

The passage of primary particles through the beam line affects the acceptance, i.e. the fraction of primary particles reaching the sensitive volume, and the characteristics of the particles entering it.

The spectrum of protons at the entrance of the sensitive volume, after traversing the beam line, is shown in Fig. 3. The proton interactions in the beam line shift the mean of the energy distribution to a lower value than the nominal beam energy of 63.95 MeV and broaden its width, originally of 300 keV: the peak part of the spectrum in Fig. 3 is well fit (with

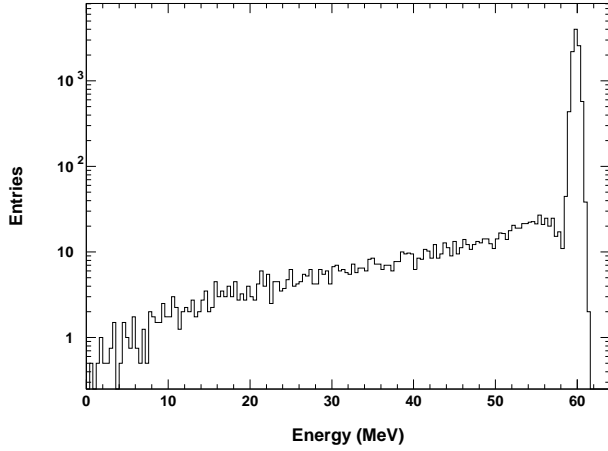


Fig. 3. Proton energy spectrum at the entrance of the sensitive water volume, after traversing the beam line; the primary beam was modelled according to a gaussian distribution, with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV.

p-value 1) by a gaussian distribution with 59.823 MeV mean and 376 keV width (standard deviation). The energy spectrum of the protons entering the sensitive volume is characterized by an extended low energy tail, which affects the plateau of the energy deposition profile in the water volume: low energy protons stop at lower depth in the water volume, producing a localized large energy deposit typical of the Bragg peak. This effect is highlighted in Fig. 4, which compares the energy deposition profile resulting from the proton spectrum shaped by the beam line to the one produced by the same spectrum, where the tail was suppressed. To suppress the tail, primary protons at the entrance of the sensitive volume, whose energy differed by more than three standard deviations from the 59.823 MeV peak value, were not tracked further. A data sample consisting of 280000 primary protons entering the sensitive volume was used for producing the energy deposition profile in Fig. 4; this sample is larger than the ones shown in other figures to better expose the effects of the low energy proton tail.

In the plateau at lower penetration depth the difference between the two curves amounts to more than 15%, while the shape of the peak is hardly affected. This feature affects parameters used in clinical practice to evaluate the quality of the irradiation, like the peak to entrance ratio. This analysis shows that imprecise knowledge of the beam line geometry and materials can affect the energy deposited in the sensitive volume; for accurate simulation of the energy deposition in the sensitive water volume, not only accurate modelling of particle interactions in water is important, but also in the materials of the beam transport line.

In experimental practice, the features of the particle spectrum should be taken into account in the choice of the optimal technique for the validation of simulation models: for instance, Faraday-cup based dosimetry is more sensitive to the energy distribution of the proton beam than ionization chambers or calorimeters [154]; the presence of a small admixture of low-energy scattered protons can lead to significant errors in absorbed dose determination with Faraday cup. [154]- [156].

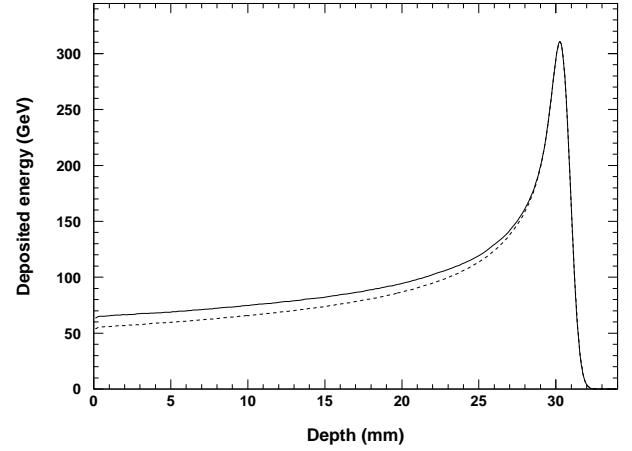


Fig. 4. Energy deposition versus penetration depth resulting from the whole spectrum of particles entering the water phantom (solid line) and produced only by the peak portion of the spectrum of Fig. 3 (dashed line), excluding the low energy tail. The peak portion of the primary proton spectrum at the entrance has 59.823 MeV mean energy and 376 keV standard deviation; the figure shows the deposited energy in each slice of the longitudinally segmented readout geometry, integrated over 280000 primary protons entering the sensitive water volume.

The acceptance values calculated with the various physics model combinations listed in Table I are all compatible within the statistical uncertainties. Further details about the effects of physics models, and their associated epistemic uncertainties, on the determination of the acceptance in the sensitive volume are examined in section V-G.

B. Water mean ionisation potential

Knowledge of the mean excitation energy of a medium is needed to calculate the energy loss of a charged particle penetrating that medium; various theoretical calculations and experimental measurements are documented in the literature concerning this parameter. The value (75 ± 3) eV recommended in ICRU Report 49 [54] is commonly used in Monte Carlo codes (e.g. Geant4, FLUKA, MCNPX). Nevertheless, values differing from this reference have recently been proposed: among them, 80.8 ± 3 eV in [157], based on theoretical and experimental considerations, 61.77 eV in [158], 79.7 ± 0.5 eV in [159] and 81.6 eV in [160]; a lower value of 67.2 eV is assumed in ICRU Report 73 [64], [157]. An experimental determination of 78.4 ± 1.0 eV was reported in [165], where a Geant4 simulation encompassing the *ICRU49* stopping power model was utilized. The lack of consensus about the value of this parameter corresponds to an epistemic uncertainty in the simulation.

The effect of the uncertainty of the water ionization potential was estimated by performing simulations with values of 75 eV (as in the reference physics configuration), 67.2 eV and 80.8 eV; apart from this feature, the application activated the physics configuration summarized in Table II. The longitudinal energy deposition profiles corresponding to different values are shown in Fig. 5. A small shift in the depth of the peak is visible; the 67.2 eV and 80.8 eV settings displace the peak to the adjacent readout geometry slices with respect to the

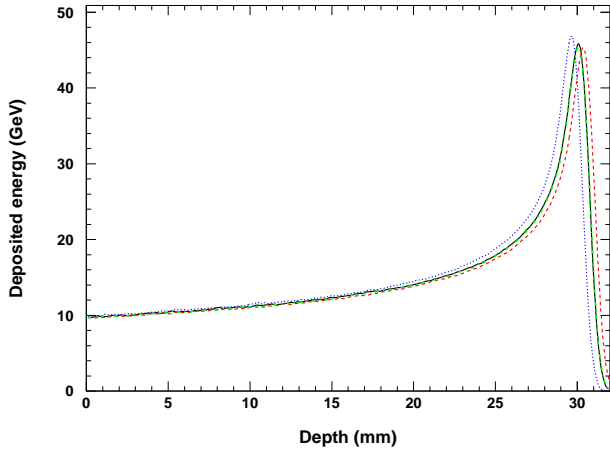


Fig. 5. Bragg peak profile resulting from different water ionisation potentials and proton beam energies: 75 eV [54] (solid black curve), 67.2 eV (dotted blue curve), 80.8 eV (dashed red curve) with proton beam energy of 63.95 MeV, and 80.8 eV (dot-dashed green curve) with proton beam energy of 63.65 MeV. The Bragg peak is from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water; the plot shows the deposited energy in each slice of the longitudinally segmented readout geometry, integrated over all the generated events.

depth resulting from the water ionization potential set at 75 eV. As described in section IV-A, the resolution in the Bragg peak location achievable in the simulation is $200 \mu\text{m}$, which corresponds to the longitudinal segmentation of the readout geometry.

Similar effects were also observed in simulations with the SHIELD-HIT code [161] and with FLUKA [162]; [35] reports approximately 1 mm shift between Geant4-based Bragg peak simulations of 85.6 and 209.2 MeV proton beams, respectively with 75 eV and 70.9 eV water mean ionization potential (however, without specifying the longitudinal resolution of the deposited energy collection).

In experimental practice, the ionization potential is usually treated as a free parameter in the simulation, which is adjusted to improve the match between experimental and simulated data (e.g. [163], [164], [166], [167]). It is worth noting that different optimal values of this parameter were identified in the literature to best match the respective experimental data.

The experimental environment typical of therapeutic beam lines provides limited insight into this simulation feature, due to the common practice of empirically determining the optimal beam parameters based on the comparison between simulated and observed depth dose profiles, as discussed in section I. A test was performed to investigate this issue: two simulations were executed with different, but equally plausible settings - respectively with 63.95 proton beam energy and 75 eV ionization potential, and with 63.65 MeV proton beam energy and 80.8 eV ionization potential; the beam energy shift is compatible with typical uncertainties of the experimental environment under study. The resulting longitudinal energy deposition profiles, shown in Fig. 5, are practically undistinguishable: the peak positions coincide, and the goodness of fit tests mentioned in section IV-C confirm their compatibility at 90% confidence level. The comparison with experimental

data, whose beam energy is not known a priori with adequate precision, would not be capable of discriminating the accuracy of such distributions.

The systematic effects highlighted by this analysis are relevant only when the simulation is expected to play a predictive role. In common applications, where the simulation is used only for verification purpose, the empirical adjustments of the water mean ionization potential and of the proton beam parameters mask any potential systematic effects. The experimental discrimination of the simulation accuracy resulting from different water ionization potential values would require precise knowledge of the beam parameters and accurate measurements of the energy deposition as a function of depth, such that displacements of the Bragg peak smaller than $200 \mu\text{m}$, associated with the value of this parameter, could be appreciated.

C. Proton stopping powers

Compilations of proton stopping powers are available in [54], [55], [56], and in the SRIM [168] code. Despite the wide body of experimental data, theoretical calculations and empirical models of proton stopping powers, no consensus has yet been achieved on definitely established values. Evaluations of empirical and theoretical stopping power models reported in the literature [169], [170] are limited to a few elements and compounds; they highlight differences among the various compilations. According to these analyses, more recent stopping power models do not necessarily correspond to improved accuracy; some models describe the stopping powers for some materials well, but appear less accurate for other materials.

Due to this controversial situation, proton stopping powers are a source of epistemic uncertainty in the simulation results. A study was performed to evaluate the effects on the Bragg peak profile related to different stopping power models implemented in Geant4.

The simulations were performed with physics settings as in Table II, apart from configuring the low energy hadron ionization process with various stopping power parameterisation models. The energy range of application was set according to the recommendations of the respective references for the ICRU49, Ziegler77 and Ziegler85 parameterisation models; lacking specific documentation about the applicability of the Ziegler2000 parameterisation, this model was applied up to 2 MeV, as for ICRU49.

The results are illustrated in Fig. 6: the various proton stopping power models produce slightly different energy deposition profiles; the Ziegler77 and Ziegler85 models produce almost identical results. The peaks associated with alternative proton stopping power models are located in adjacent longitudinal readout geometry slices with respect to the peak produced by the reference configuration of Table II, including ICRU49 stopping powers; as described in section IV-A, the longitudinal readout segmentation is $200 \mu\text{m}$. Apart from the shift in the peak position, the shapes of the energy deposition profiles are statistically compatible at 90% confidence level according to all the goodness of fit tests mentioned in section IV-C.

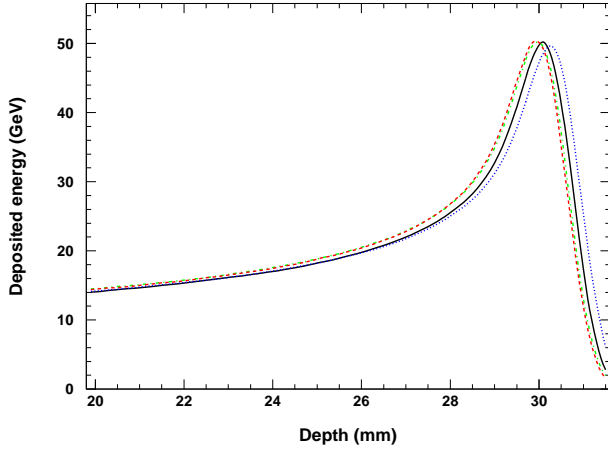


Fig. 6. Energy deposition as a function of depth resulting from different proton stopping power models: *ICRU49* (solid black curve), *Ziegler77* (dash-dotted green curve), *Ziegler85* (dashed red curve) and *Ziegler2000* (dotted blue curve); the profiles correspond to the *Ziegler77* and *Ziegler85* models are barely distinguishable in the plot, due to the great similarity of their simulation results. The Bragg peak is from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water; the plot shows the deposited energy in each slice of the longitudinally segmented readout geometry, integrated over all the generated events.

Similarly to the discussion in the previous section concerning the water mean ionization potential, the epistemic uncertainty related to the stopping power model used in the simulation turns into systematic effects only when the simulation is required to play a predictive role; otherwise, as shown in the previous case, a small adjustment of the proton beam energy, compatible with typical experimental uncertainties, would shift the energy deposition profiles deriving from different stopping power models into statistically equivalent distributions. These considerations suggest that typical proton therapy experimental environments would not be sensitive to the differences of stopping power models, nor would they be adequate to discriminate their accuracy. All the available stopping power models appear equally suitable to that simulation environment; in this respect, one can observe that the use of different Geant4 models has been reported with satisfactory agreement against experimental data, despite the fact that they determine different Bragg peak depths: for instance, *Ziegler2000* in [31] and *ICRU49* in [27], [30], [33].

If predictive capabilities are required from the simulation, higher precision experimental measurements would be necessary to discriminate the accuracy of the existing models with the capability of appreciating shifts in the peak depth smaller than $200 \mu\text{m}$.

This context should be taken into account when considering comparative evaluations of the accuracy of simulation models: the procedure of empirically adjusting the parameters in the simulation, based on a selected physics configuration, to best fit the experimental data is prone to bias further comparisons of other simulation models with the same data. The estimate of the relative accuracy of alternative physics models would require the capability of comparing the simulation outcome to measurements, without privileging any of the models to

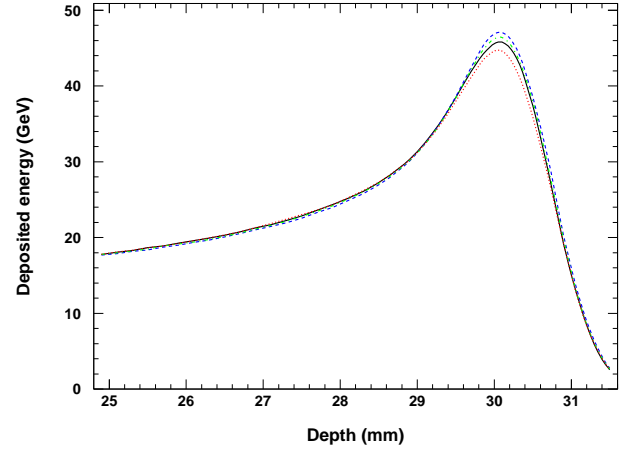


Fig. 7. Energy deposition profiles associated with various proton elastic scattering options: "U-elastic" (solid black curve), Bertini (blue dashed curve), LEP (green dot-dashed curve) and CHIPS (red dotted curve). The Bragg peaks are from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water; the plot shows the deposited energy in each slice of the longitudinally segmented readout geometry, integrated over all the generated events.

constrain any simulation parameters.

D. Hadronic elastic scattering

Four elastic scattering modeling options available in Geant4 were compared: the *G4UHadronElasticProcess* with the *G4HadronElastic* model, the *G4HadronElasticProcess* process with the *G4LElastic* (LEP) or *G4ElasticCascadeInterface* (Bertini elastic) models, and the CHIPS *G4QElastic* process.

The simulation application activated the physics configuration as in Table II for all other features apart from proton elastic scattering. The longitudinal energy deposition profiles resulting from the various simulation configurations are shown in Fig. 7. The distribution of the relative difference of the energy deposited in each longitudinal slice of the sensitive volume with respect to the outcome from the reference configuration of Table II is shown in Fig. 8 for the various options; the differences are mostly comprised within $\pm 2\%$.

The compatibility of the energy deposition profiles associated with the various elastic scattering options is confirmed by goodness-of-fit tests, whose results are reported in Table III. All the tests fail to reject the null hypothesis of compatibility with the profile deriving from the reference configuration, with 0.1 significance. A subset of elastic scattering modeling options, limited to the "U-elastic", "Bertini" and "LEP" ones, was compared in the context of Geant4 8.1p02 and 9.1 versions as well. All the considered elastic scattering options were compatible with 0.1 significance within a given Geant4 version; the results concerning the comparison of the left branch of the energy deposition profiles are reported in Table III.

If the differences between the outcome of two simulation configurations were due only to statistical fluctuations, one would expect them to be distributed, as a function of depth, in a rather large number of short sequences (runs) of consecutive positive and negative values; the Wald-Wolfowitz test [171]

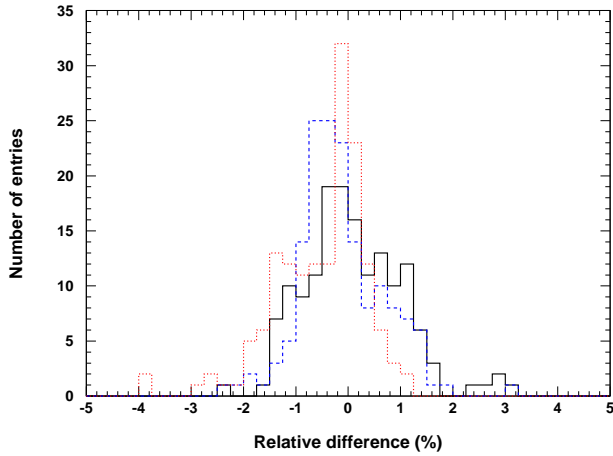


Fig. 8. Relative difference of the energy deposited in the longitudinal slices of the sensitive water volume for various elastic scattering configurations, with respect to the results of the configuration as in Table II with *G4UHadronElasticProcess* and the *G4HadronElastic* model: Bertini (solid black histogram), LEP (blue dashed histogram) and CHIPS (red dotted histogram) elastic scattering. The relative difference is calculated in each slice of the longitudinal segmentation of the readout geometry associated with the sensitive volume. The energy deposition derives from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water.

was applied to evaluate this hypothesis. The resulting p-value is smaller than 0.001 for all the test cases; therefore one can infer some systematic effects associated with the choice of the elastic scattering model in the simulation. It is worth remarking that the conclusion drawn from the Wald-Wolfowitz test does not contradict the result of the goodness-of-fit tests: the two types of tests, respectively evaluating the differences between two distributions in terms of sign and of distance, are complementary. A feature of the energy deposition profiles hinting at systematic differences is visible in the vicinity of the Bragg peak in Fig. 7, where alternative elastic scattering options appear associated with sequences of energy deposition consistently larger or smaller than those deriving from the configuration of Table II encompassing the "U-elastic" elastic scattering model.

For the use case under study, the small differences exhibited by the various simulation models look compatible with the experimental resolution typical of the application domain (of the order of 2-2.5% [154]); therefore, the peculiarities of the models do not affect the outcome of the simulation significantly. Based on these results, one can conclude that at the present stage all the Geant4 elastic scattering options are equivalent for the use case considered in this study. Validation against experimental data concerning the energy range and target materials pertinent to this use case would strengthen the predictive reliability of the simulation.

E. Hadronic inelastic cross sections

The proton total inelastic cross sections are an important parameter in the simulation of therapeutical proton beams, since they determine the amount of proton loss from the primary beam.

TABLE III
P-VALUE OF GOODNESS-OF-FIT TESTS COMPARING HADRONIC ELASTIC SCATTERING OPTIONS

| Version | Range | Model | Kolmogorov Smirnov | Anderson Darling | Cramer von Mises |
|---------|--------------|---------|--------------------|------------------|------------------|
| 9.3 | Whole | Bertini | 1 | 1 | 1 |
| | | LEP | 1 | 1 | 1 |
| | | CHIPS | 0.997 | 0.997 | 0.999 |
| | Left branch | Bertini | 1 | 1 | 1 |
| | | LEP | 1 | 1 | 1 |
| | | CHIPS | 0.996 | 0.982 | 0.999 |
| | Right branch | Bertini | 1 | 0.986 | 0.972 |
| | | LEP | 1 | 0.986 | 0.972 |
| | | CHIPS | 1 | 0.986 | 0.972 |
| 9.1 | Left branch | Bertini | 1 | 1 | 1 |
| | | LEP | 0.996 | 0.989 | 1 |
| 8.1 | Left branch | Bertini | 1 | 0.999 | 0.997 |
| | | LEP | 1 | 1 | 1 |

Two configurations of cross sections were evaluated in this study: those implemented in *G4HadronInelasticDataSet*, originating from GHEISHA, and those implemented in *G4ProtonInelasticCrossSection* and *G4NeutronInelasticCrossSection*, respectively for protons and neutrons, covering the energy range above 6.8 MeV. Apart from this feature, all the other physics options in the simulation were set as in Table II.

Both cross sections derive from parameterisations of experimental data; it is not clear whether the comparisons available in the literature concern the calibration of the parameterisation with experimental data, or represent the cross section model validation.

No significant dependence on the cross section options is observed regarding the proton acceptance in the sensitive water volume, which is affected by the interactions to which primary protons are subject in the materials of the beam line.

The two sets of cross sections determine some difference in the occurrence of the proton inelastic scattering process associated with them in the sensitive water volume. Confidence intervals for this quantity were calculated, using Student's *t* distribution, based on the simulation sample activating the cross sections as in [94]. The 99% confidence interval for the mean value of hadronic inelastic scattering occurrences lies between 1688 and 1849, when one million primary protons are generated, while the number of occurrences with the GHEISHA-like cross section data set is 1654; this value is significantly different from the mean number of inelastic scattering occurrences determined by the cross sections of [94].

Nevertheless, the effect of this difference on the longitudinal energy deposition appears negligible. The distribution of the relative differences of the energy deposition profiles associated with the two options is shown in Fig. 9; it is consistent with typical experimental uncertainties in hadron therapy practice. The longitudinal energy deposition profiles resulting from the two cross section options, with other physics settings as in Table II, are compatible at 90% confidence level according to all the aforementioned goodness-of-fit-tests.

Therefore one can conclude that the characteristics of the two hadronic cross section data sets do not affect the simula-

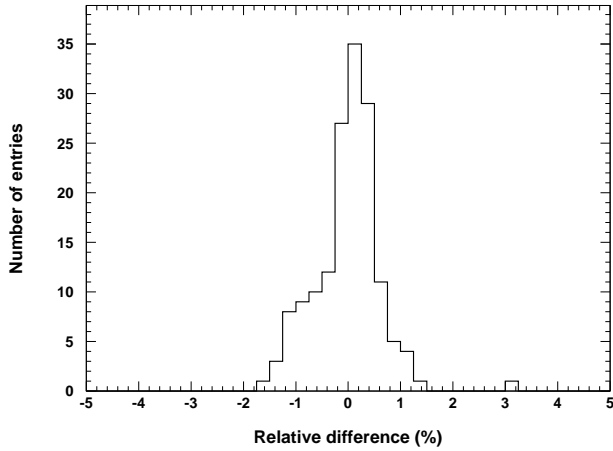


Fig. 9. Percent relative difference of the energy deposition profiles resulting from the activation of *G4HadronInelasticDataSet* hadronic inelastic cross sections or *G4ProtonInelasticCrossSection* and *G4NeutronInelasticCrossSection*; all the other simulation options are identical in the two cases, and set as in Table II. The energy deposition derives from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water. The relative difference is calculated in each slice of the longitudinal segmentation of the readout geometry associated with the sensitive volume.

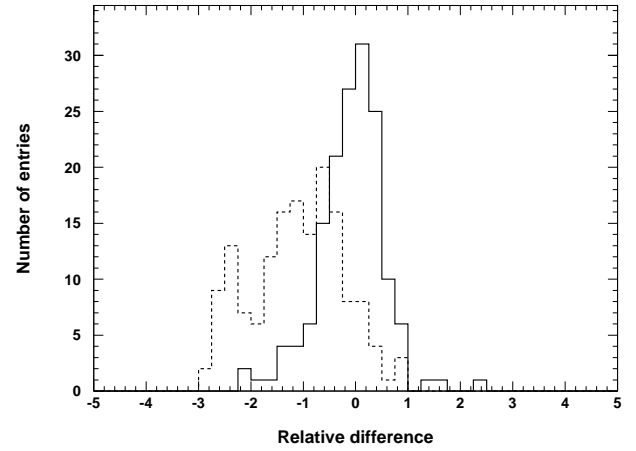


Fig. 10. Relative difference of longitudinal energy deposition profiles associated with the GEM evaporation model, with respect to the “reference physics configuration”: simulation based on Geant4 9.3 (solid line) and 9.1 (dashed line). The observed systematic effect is related to the correction of a software feature in Geant4 9.3. The energy deposition derives from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water. The relative difference is calculated in each slice of the longitudinal segmentation of the readout geometry associated with the sensitive volume.

tion of the proton depth dose profiles in the use case considered in this study.

F. Hadronic inelastic scattering models

Several alternative hadronic inelastic scattering modeling options were evaluated: the Precompound model, the Bertini and Liège cascade models, the LEP parameterised model and the CHIPS model. In addition, a few configuration options of the nuclear deexcitation phase, accessible through the interface of the *G4ExcitationHandler* class instantiated by the Precompound model, were evaluated together with the Precompound model: the generalized evaporation (GEM) model replacing the default evaporation one, and the activation of the Fermi break-up. The Precompound model was evaluated both as a standalone model and as invoked by the Binary cascade model.

Proton and neutron interactions were handled consistently in each simulation configuration by activating the same hadronic inelastic model option for both particles; all the other physics features were set as in Table II.

The longitudinal energy deposition profiles produced by the various hadronic inelastic models in Geant4 9.3 appear visually undistinguishable; therefore no related figure is shown in this paper.

The energy deposition profile produced with the GEM evaporation model closely resembles the one deriving from the default evaporation model instantiated in connection with the Precompound model, as seen in Fig. 10; this observation is confirmed by the results of the goodness-of-fit tests in Table IV with 0.1 significance. Differences related to the use of the two models were visible with previous Geant4 versions, as shown in Fig. 10; the GEM implementation was subject to improvements in Geant4 9.3 [73].

The distributions of the secondary particles produced in association with the two evaporation models look consistent,

compatible with statistical fluctuations; the secondary proton distributions are shown in Fig. 11 as an example. The lack of visible effects does not necessarily mean that these two models are characterized by identical features; rather, it shows that the use case under study is not sensitive to their possible difference.

From this analysis one can conclude that the evaporation model options are equivalent for the simulation of the longitudinal energy deposition; as documented in section VI, the GEM model is computationally faster than the default one in the application under study.

The evaporation process of nuclear deexcitation is based on the hypothesis that the excitation energy is high and approximately equally distributed among the nucleons: this assumption is justified for heavy nuclei, but it is not applicable to the water target considered in this use case. It is generally accepted that the Fermi break-up represents a more appropriate theoretical description of the nuclear deexcitation process for light nuclei: in MCNPX and FLUKA the Fermi break-up is applied to nuclei with atomic mass up to 17, whereas evaporation is applied to heavier nuclei; in Geant4 it is not invoked by default in the deexcitation of light nuclei, although the public interface of *G4ExcitationHandler* allows users to modify the default settings.

The effects of the Fermi break-up were evaluated by activating it in the simulation for nuclei with atomic number smaller than 10; they are visible in the spectrum of the produced secondaries, with respect to those produced by nuclear deexcitation proceeding through evaporation. From a theoretical perspective, the application of an evaporation model to the deexcitation of light nuclei is expected to overestimate the production of secondary protons in the lower energy range; this effect is indeed observed in Fig. 11. The activation of the Fermi break up affects the longitudinal energy deposition, with

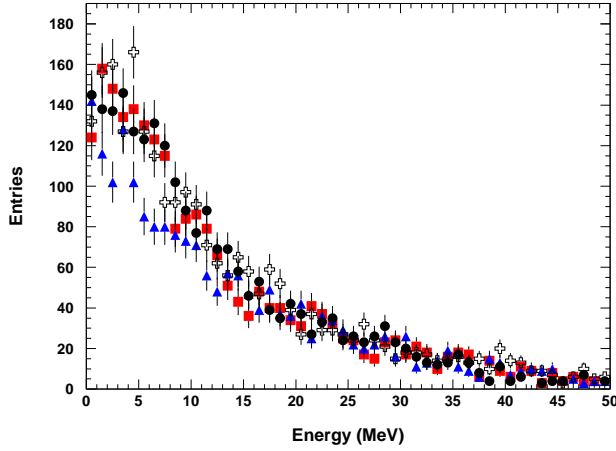


Fig. 11. Energy spectrum of secondary protons produced with different configurations of the Precompound model: default configuration as in Table II (black circles), configuration with GEM evaporation (red squares), configuration activating Fermi break up (blue triangles) and configuration activating the Binary Cascade model (white crosses), which in turn invokes the Precompound model to handle the preequilibrium phase. The secondary spectra derive from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water.

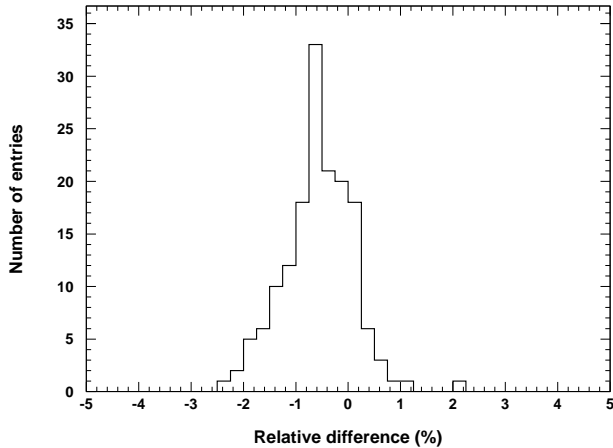


Fig. 12. Relative difference of the longitudinal energy deposition profile deriving from the activation of the Fermi break-up, with respect to the profile deriving from the default configuration of the Precompound model; all the other settings are as in Table II. The energy deposition derives from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water. The relative difference is calculated in each slice of the longitudinal segmentation of the readout geometry associated with the sensitive volume.

respect to that resulting from the default nuclear deexcitation settings: their relative difference, shown in Fig. 12, exhibits an asymmetric distribution shifted towards negative values. This effect hints at a systematic contribution of the Fermi break-up to decrease the longitudinal energy deposition; nevertheless, the observed differences are consistent with typical uncertainties in experimental proton therapy practice.

A similar asymmetry in the longitudinal energy deposition difference is observed when the preequilibrium phase is associated with intranuclear transport; this effect is shown in Fig. 13, which concerns two simulations involving the Precompound model, respectively as an independent model and invoked

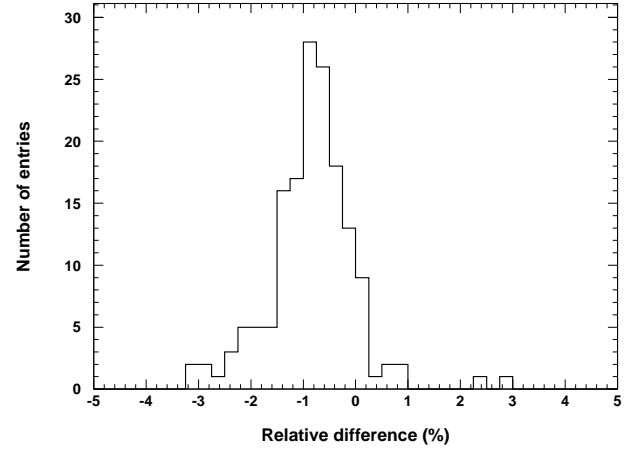


Fig. 13. Relative difference of the longitudinal energy deposition profile deriving the activation of the Precompound model either as a standalone model, or as invoked by the Binary cascade model. The energy deposition derives from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water. The relative difference is calculated in each slice of the longitudinal segmentation of the readout geometry associated with the sensitive water volume.

through the Binary Cascade model. The transition between the cascade process of intranuclear transport and the preequilibrium is determined by empirical considerations [103], which are specific to each software implementation: for instance, in Geant4 Binary cascade model cascading continues as long as there are particles above a 70 MeV kinetic energy threshold [103] (along with other conditions required by the algorithm), while a smooth transition around 50 MeV is implemented in FLUKA [10]. The empirical features of the algorithm correspond to lack of knowledge from physical principles to determine the transition between the two régimes; the analysis shows that this epistemic uncertainty can be a source of systematic effects, such as the bias of the distribution in Fig. 13. This effect, which is of a magnitude comparable to typical experimental uncertainties in hadron therapy measurements, could be significant in use cases where precise predictive power is expected from the Monte Carlo simulation.

No such asymmetries, with respect to simulating the preequilibrium phase only (as in the Precompound model), are observed with two other configurations involving intranuclear cascade models - the Bertini and Liège ones. It is worth remarking that the Liège model does not involve a preequilibrium phase at all, while the Bertini cascade model does. The apparent absence of consistent trends related to the adopted physical approach (modeling intranuclear cascade, preequilibrium and their interplay) suggests that the observed behavior of the code may be influenced by other implementation details on top of the basic physics modeling approach; this consideration adds further complexity to the effort of identifying the sources of epistemic uncertainty in the simulation, which is a necessary step towards their reduction or their control.

The relative differences of the energy deposition profiles concerning other hadronic inelastic models with respect to the outcome deriving from the Precompound one are illustrated in Fig. 14.

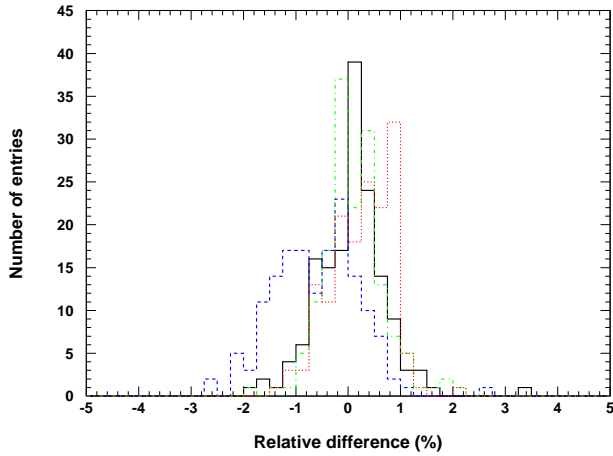


Fig. 14. Relative difference of the energy deposited in the longitudinal slices of the sensitive water volume associated with various hadronic inelastic options, with respect to the configuration of Table II encompassing the Precompound model: Bertini (solid black histogram), LEP (blue dashed histogram), Liège (green dot-dashed histogram) and CHIPS (red dotted histogram). The energy deposition derives from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water.

All the final state models of hadronic inelastic scattering produce statistically compatible results at 90% confidence level in the considered use case, as shown in Table IV.

The Wald-Wolfowitz test concerning the difference with respect to the reference physics configuration results in a p-value smaller than 0.001 for all the considered modeling options, with the exception of the comparison involving the Liège cascade model, for which the p-value is 0.360. These results suggest the presence of some systematic effects due to the choice of the hadronic inelastic models in the simulation; some asymmetries and bias with respect to zero are indeed visible in the distributions in Fig. 14, namely the one concerning the LEP inelastic model, apart from the previously discussed effects in Fig. 12 and 13.

Like the results discussed in section V-D, this result suggests that the selection of the hadronic inelastic model activated in the simulation can be source of systematic effects. Nevertheless, the differences concerning the longitudinal energy deposition patterns appear compatible with typical experimental uncertainties in proton therapy dosimetry; therefore the systematic effects identified by the Wald-Wolfowitz test would be negligible in that software application context. They could become relevant in use cases where higher accuracy is demanded.

The implementation of the Geant4 Precompound model was subject to improvements [38] in Geant4 9.2. Nevertheless, these modifications do not appear to have affected the features of the longitudinal energy deposition pattern significantly, since the goodness-of-fit tests in Table IV show that the energy deposition profiles associated with this model were compatible with those deriving from other hadronic inelastic models in previous Geant4 versions, as well as in the 9.3 one. This remark is relevant to previous applications of the Precompound model to the use case under study, which are archived in the literature.

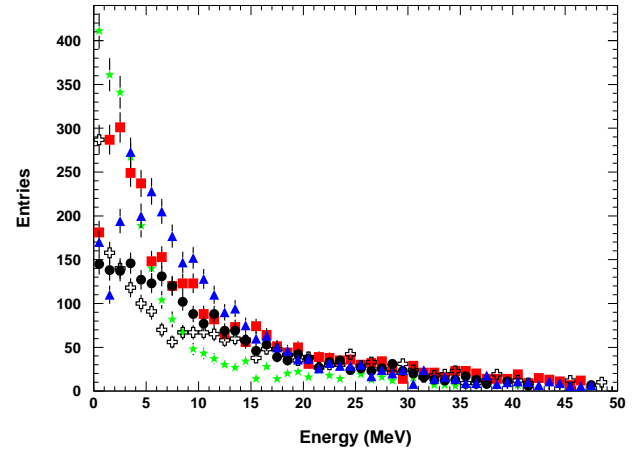


Fig. 15. Secondary proton spectrum resulting from various hadronic inelastic models: Precompound (black circles), Bertini (red squares), LEP (white crosses), Liège (blue triangles) and CHIPS (green stars). The secondary particle spectrum derives from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water.

Nevertheless, despite their similarity at determining the longitudinal energy deposition profile, some hadronic inelastic models exhibit very different characteristics regarding the secondary particles they generate: the secondary proton, neutron and α particle spectra are shown in Fig. 15 to 17. Radiotherapy applications can be affected by secondary particles within the target volume and outside it, both laterally and beyond the distal edge of the Bragg peak [155]; concerns for the risks due to secondary particles in proton therapy are discussed in the literature [173]. The analysis documented in the previous paragraphs shows that the different secondary particle production patterns do not produce significant effects on the longitudinal energy deposition profile; the quantification of possible effects on the lateral energy deposition pattern would require substantially larger data samples, which were not achievable with the limited computational resources available to the authors in the course of the project documented in this publication, and is outside the scope of this paper.

The identification of actual systematic effects related to hadronic inelastic models, and their quantitative estimate, would require experimental measurements with adequate accuracy to discriminate not only the features of the energy deposition distribution, but also the characteristics of the secondary particles they produce.

G. Multiple Coulomb scattering

The configuration of proton multiple scattering simulation in a Geant4 application involves the selection of the multiple scattering process and models to be activated, and setting some parameters used by the multiple scattering algorithm. Default options are provided in Geant4 kernel for the model and parameters associated to multiple scattering processes; they are summarized in Table V for a set of recent Geant4 versions.

The analysis evaluated whether recent evolutions of the code between Geant4 8.1 and 9.3 versions, which involve

TABLE IV
P-VALUE OF GOODNESS-OF-FIT TESTS COMPARING HADRONIC INELASTIC SCATTERING OPTIONS

| Geant4 version | Test range | Hadronic model | Kolmogorov Smirnov | Anderson Darling | Cramer von Mises |
|----------------|--------------|----------------|--------------------|------------------|------------------|
| 9.3 | Whole | Bertini | 1 | 1 | 1 |
| | | LEP | 0.954 | 0.988 | 0.984 |
| | | Liège | 1 | 1 | 1 |
| | | CHIPS | 1 | 1 | 1 |
| | | GEM | 1 | 1 | 1 |
| | | Fermi break-up | 1 | 1 | 1 |
| | | Binary | 0.954 | 0.938 | 0.973 |
| | Left branch | Bertini | 1 | 1 | 1 |
| | | LEP | 0.945 | 0.961 | 0.979 |
| | | Liège | 1 | 1 | 1 |
| | | CHIPS | 1 | 1 | 1 |
| | | GEM | 1 | 1 | 1 |
| | | Fermi break-up | 1 | 1 | 1 |
| | | Binary | 0.945 | 0.858 | 0.962 |
| | Right branch | Bertini | 1 | 0.986 | 0.972 |
| | | LEP | 1 | 0.986 | 0.972 |
| | | Liège | 1 | 0.986 | 0.972 |
| | | CHIPS | 1 | 0.986 | 0.972 |
| | | GEM | 1 | 0.986 | 0.972 |
| | | Fermi break-up | 1 | 0.986 | 0.972 |
| | | Binary | 1 | 0.986 | 0.972 |
| 9.1 | Left branch | Bertini | 0.981 | 0.901 | 0.980 |
| | | LEP | 0.945 | 0.949 | 0.937 |
| 8.1 | Left branch | Bertini | 1 | 1 | 1 |
| | | LEP | 0.996 | 0.814 | 0.847 |

TABLE V
DEFAULT SETTINGS OF RELEVANT MULTIPLE SCATTERING PROCESSES

| Geant4 version | Process | Range Factor | Step Limit | Lateral Displacement | skin | Geometry Factor | Model |
|----------------|-----------------------|--------------|------------|----------------------|------|-----------------|---------|
| 8.1 | G4MultipleScattering | 0.02 | 1 | | | | Urban |
| 9.1 | G4MultipleScattering | 0.02 | 1 | 1 | 0 | 2.5 | Urban |
| 9.2p0.3 | G4MultipleScattering | 0.02 | 1 | 1 | 3 | 2.5 | Urban |
| 9.3 | G4MultipleScattering | 0.04 | 1 | 1 | 3 | 2.5 | Urban92 |
| 9.3 | G4hMultipleScattering | 0.2 | 0 | 1 | 3 | 2.5 | Urban90 |

different model and parameter settings, could be the source of systematic effects in Geant4-based simulations for proton therapy applications, originating from epistemic uncertainties in the simulation model. Two issues were addressed: the effects related to different multiple scattering processes, *G4MultipleScattering* and *G4hMultipleScattering*, and those due to changes in the *G4MultipleScattering* process since the 8.1 release. It should be remarked that in the following analysis the behaviour associated with multiple scattering may result not only directly from the implementation of the two above mentioned classes, but also from behavior inherited from their base classes or acquired through aggregation of, or dependence on, other classes, as determined by the software design of Geant4 multiple scattering domain. The results are reported for Geant4 versions 8.1p02, 9.1, 9.2p03, and 9.3.

Two multiple scattering processes, *G4hMultipleScattering* and *G4MultipleScattering*, are applicable to protons in Geant4 9.3. The former was first introduced in Geant4 8.2 to provide faster simulation of hadron transport; the latter complies with an earlier class interface and is planned to be withdrawn from later releases. In Geant4 9.3 the common base class *G4VMultipleScattering* accounts for public member functions formerly specific to *G4MultipleScattering*. *G4hMultipleScattering* can be configured to acquire equiva-

lent behavior to *G4MultipleScattering* by applying the same settings (model and parameters) as in *G4MultipleScattering* listed in Table V. For convenience, the configuration of *G4hMultipleScattering* equivalent to *G4MultipleScattering* is still indicated as *G4MultipleScattering* in the following.

Only the default settings listed in Table V were tested; the large effects related to epistemic uncertainties observed in this limited interval analysis, which are documented in the following, suggest that this complex problem domain would benefit from a larger-scale dedicated study beyond the scope of this paper.

To acquire sound evidence of effects associated with multiple scattering settings, the comparisons were performed over five physics configurations: the set of processes and models (apart from multiple scattering) as in Table II, and variants of it consisting of “LEP” and “Bertini” inelastic scattering together with “U-elastic” elastic scattering, and “LEP” and “Bertini” elastic scattering together with the Precompound hadronic inelastic model. Common effects observed in such an extended set of test cases could be reasonably associated with the multiple scattering domain, excluding their possible origin from intrinsic features of a single physics configuration.

The resulting longitudinal energy deposition distributions associated with proton multiple scattering options are shown

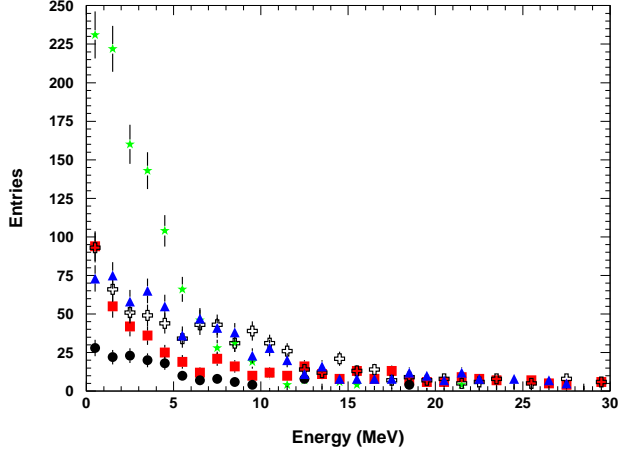


Fig. 16. Secondary neutron spectrum resulting from various hadronic inelastic models: Precompound (black circles), Bertini (red squares), LEP (white crosses), Liège (blue triangles) and CHIPS (green stars). The secondary particle spectrum derives from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water.

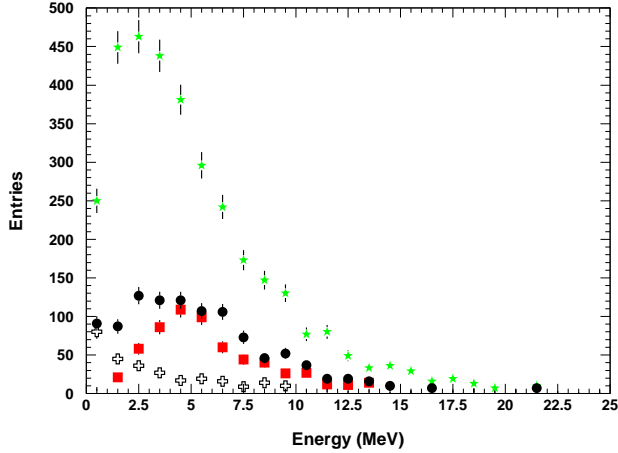


Fig. 17. Secondary α particle spectrum resulting from various hadronic inelastic models: Precompound (black circles), Bertini (red squares), LEP (white crosses) and CHIPS (green stars); no α particles appear to be produced by the Geant4 implementation of the Liège model. The secondary particle spectrum derives from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water.

in Fig. 18. The two multiple scattering processes produce visibly different longitudinal energy deposition profiles; the extent of the differences can be quantitatively appraised in Fig. 19, which shows the variation of the longitudinal energy deposition profiles simulated with *G4hMultipleScattering* in Geant4 9.3 with respect to equivalent simulations performed with *G4MultipleScattering* in Geant4 9.3, 9.1 and 8.1p02. The plot shows results produced with the Bertini hadronic inelastic option; similar results are obtained with the other physics configurations subject to comparison. The energy deposition profile of Geant4 9.2p03 is not shown in Fig. 19 to avoid clogging the plot with many curves; it lies in between the profiles produced by Geant4 9.3 with *G4hMultipleScattering* and Geant4 9.1.

The results of goodness-of-fit tests comparing longitudinal

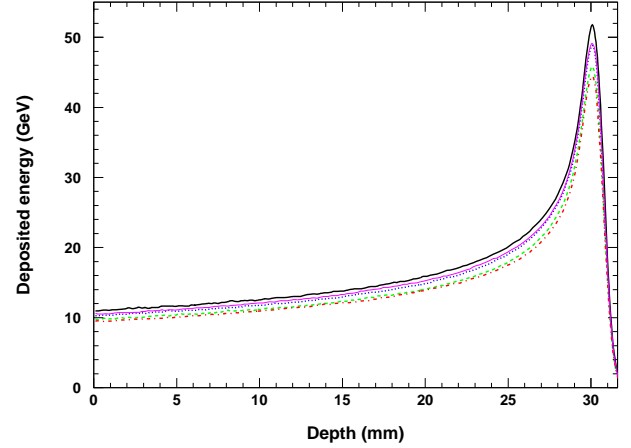


Fig. 18. Bragg peak profile resulting from different multiple scattering processes and Geant4 versions: *G4hMultipleScattering* (black, thick solid line) in Geant4 9.3, *G4MultipleScattering* in Geant4 9.3 (dashed, green line), Geant4 9.2p03 (pink, thin solid line), Geant4 9.1 (dotted, blue line) and Geant4 8.1p02 (dash-dotted, red line). The same physics configuration was activated in all the simulations, apart from the multiple scattering setting. The Bragg peak is from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water; the plot shows the deposited energy in each slice of the longitudinally segmented readout geometry, integrated over all the generated events.

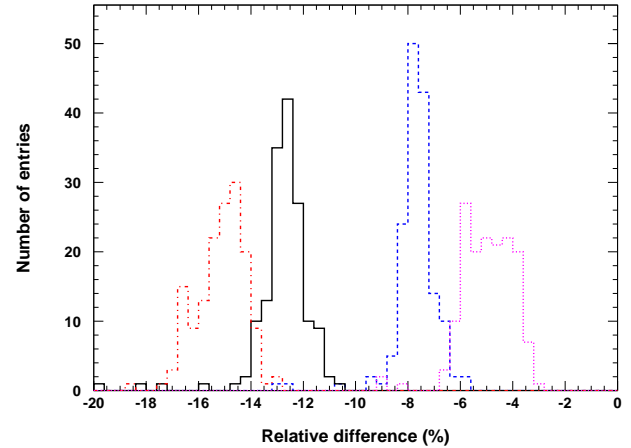


Fig. 19. Relative difference of the energy deposited in the longitudinal slices of the sensitive water volume associated with different multiple scattering processes and Geant4 versions; the difference is calculated in each longitudinal slice of the readout geometry with respect to a reference configuration with *G4hMultipleScattering* in Geant4 9.3 for identical configurations with *G4MultipleScattering* in Geant4 9.3 (solid black histogram), 9.1 (dashed blue histogram), 9.2p03 (dotted pink histogram) and 8.1p02 (dash-dotted red histogram). The reference configuration is as in Table II, except for the hadronic inelastic scattering option (Bertini instead of Precompound); this replacement is due to the greater stability of the Bertini code across the various Geant4 versions, nevertheless all other physics configurations produce similar results. The energy deposition derives from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water.

energy deposition distributions associated with either proton multiple scattering process are summarized in Table VI. The longitudinal energy deposition distributions associated with *G4hMultipleScattering* are incompatible at 99.9% confidence level with those produced by Geant4 versions 8.1p02 and 9.3 with *G4MultipleScattering*, and, apart from one test case involving the Kolmogorov-Smirnov test, at 99% confidence level with those produced by Geant4 9.1. Regarding the comparison with the profiles generated with Geant4 9.2p03, the Anderson-Darling test rejects the hypothesis of compatibility with the profiles produced with *G4hMultipleScattering* at 95% confidence level, the Kolmogorov-Smirnov does not reject it, while the Cramer-von Mises rejects it in two physics configurations and does not reject it in the other three configuration. The Anderson-Darling and Cramer-von Mises tests are considered more powerful than the Kolmogorov-Smirnov test; the Anderson-Darling test is especially sensitive to fat tails [176]. *G4MultipleScattering* was responsible of the multiple scattering process in these earlier Geant4 versions.

It is worth mentioning that [35] shows comparisons of experimental and simulated proton energy deposition profiles normalized to the number of protons in the beam; the reported simulations were performed with Geant4 8.1p01 version. The small plots in logarithmic scale and the limitation of the comparisons to qualitative appraisal prevent the reader from understanding whether the different behaviour of *G4hMultipleScattering*, and of the *G4MultipleScattering* class in later Geant4 versions, with respect to the multiple scattering implementation of Geant4 8.1, would affect the compatibility with the experimental data of [35].

Visible differences are also observed in Fig. 20 concerning the energy deposition profiles associated with *G4MultipleScattering* settings over the various versions. The total energy deposition shown in Fig. 21 exhibits evident differences associated with the various settings.

The Geant4 Low Energy Electromagnetic package, used in all the simulations, was subject to configuration and Change Management discipline [174] based on the Unified Software Development Process [175] framework until Geant4 release 9.1; the adopted software process ensured that the software of this package relevant to the use case under study did not undergo modifications between the 8.1p02 and 9.1 production versions, which could alter its physical behaviour. The same implementations of the low energy electromagnetic processes were used in the simulations based on Geant4 9.2 and 9.3; therefore, it is plausible that variations observed in the simulation productions based on different Geant4 versions are associated with evolutions in other Geant4 domains. The extent of the differences observed when comparing two Geant4 versions appears to be approximately the same over all the hadronic physics configurations activated in the simulation; since the occurrence of coherent modifications to all the Geant4 hadronic elastic and inelastic scattering implementations is not likely, this observation suggests that coherent differences would derive from modifications either to Geant4 transport kernel or to the multiple scattering domain, which are common to all the simulations. Major changes to Geant4 kernel have not been documented over the considered versions;

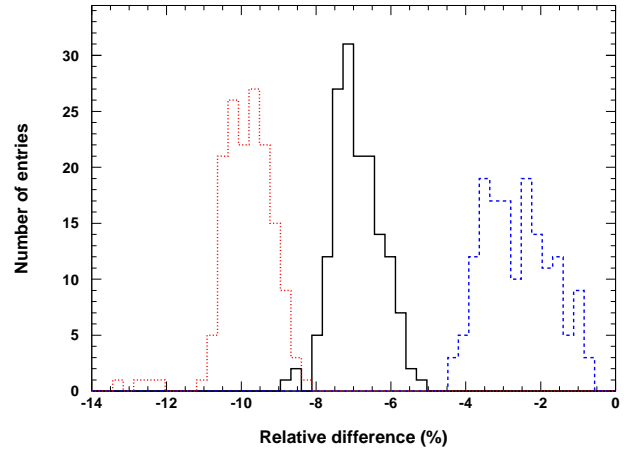


Fig. 20. Relative difference of the energy deposited in the longitudinal slices of the sensitive water volume associated with the *G4MultipleScattering* in Geant4 9.2p03, with respect to identical simulation settings in Geant4 9.3 (solid black histogram), Geant4 9.1 (dashed blue histogram), and Geant4 8.1p02 (dotted red histogram). The reference configuration is as in Table II, except for the hadronic inelastic scattering option (Bertini instead of Precompound); this replacement is due to the greater stability of the Bertini code across the various Geant4 versions, nevertheless all other physics configurations produce similar results. The energy deposition derives from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water.

the multiple scattering domain, which was subject to evolution, appears the most likely source of the observed discrepancies. Their origin is probably from epistemic uncertainties in the simulation models, whose validation in the energy range relevant to this use case is scarcely documented in literature.

The differences concerning multiple scattering settings in the various Geant4 versions are significant. The 99.9% and 99% confidence intervals for the mean value of the total energy deposition deriving from *G4hMultipleScattering* in Geant4 9.3 are shown in Fig. 21; the values deriving from Geant4 versions 8.1p02, 9.1 and 9.2p03 fall outside the 99.9% confidence interval.

The results of goodness-of-fit tests are reported in Table VI. In most test cases the longitudinal energy deposition distributions produced with Geant4 9.3 are incompatible with those produced with Geant4 9.1 at 95% confidence level; in a few cases the test statistic results in p-values close to the critical region for 0.05 significance. The null hypothesis of equivalence of the distributions subject to test is not rejected, with the same significance, in the comparisons involving Geant4 9.3 and 8.1p02 versions. This quantitative result is consistent with the qualitative appraisal of Fig. 18, where the energy deposition profile deriving from Geant4 9.3 appears closer to the one produced with the earlier 8.1p02 version. The energy deposition profiles produced with Geant4 9.2p03 are incompatible with those produced with Geant4 9.3 (with *G4MultipleScattering*) and Geant4 8.1p02 with 0.05 significance, while the goodness-of-fit tests fail to reject the hypothesis of compatibility with the profiles produced with Geant4 9.1 with 0.05 significance. The longitudinal energy deposition distributions produced with Geant4 9.1 and 8.1p02 are incompatible with 0.05 significance.

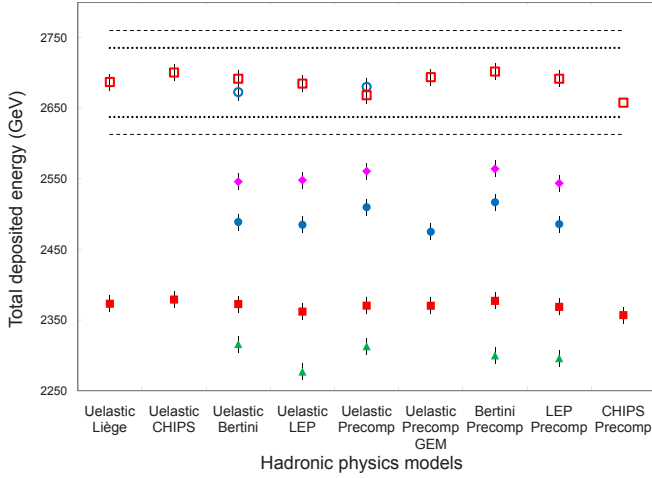


Fig. 21. Total energy deposition in the sensitive volume associated with various Geant4 versions and physics configurations: Geant4 9.3 (red squares), Geant4 9.2p03 (pink diamonds), Geant4 9.1 (blue circles) and Geant4 8.1p02 (green triangles); the filled symbols correspond to simulations activating the *G4MultipleScattering* multiple scattering process, while the empty ones correspond to the activation of *G4hMultipleScattering*. The upper and lower lines of the horizontal axis identify respectively the hadronic elastic and inelastic scattering model in each simulation configuration; the other physics options, apart from the multiple scattering under test, were as in Table II. The dashed and dotted lines represent respectively the 99.9% and 99% confidence intervals for the mean value of the total deposited energy over various Geant4 9.3 physics configurations associated with *G4hMultipleScattering*. The total energy deposition derives from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water.

The simulations with the two multiple scattering processes and with different Geant4 versions produce a significantly different total energy deposition in the sensitive volume. The results are shown in Fig. 21; the dashed and dotted lines in the plot represent respectively the 99.9% and 99% confidence intervals for the average energy deposition with *G4hMultipleScattering* over all Geant4 9.3 physics configurations.

The absolute value of the energy deposition is relevant to applications where knowledge of the actual dose released to a target is critical, like oncological treatment planning, radiation protection or radiation damage estimate. The observed differences would be significant in use cases where the simulation has a predictive role: differences greater than 10% in the dose released to a patient, like the effects observed with the various multiple scattering implementations released in Geant4, would be important in clinical applications. These use cases would not be limited to the bio-medical application domain; for instance, the use of Monte Carlo simulation to study the damage to electronic components exposed to radiation would require precise estimate of the released dose.

The average energy deposition per proton in the sensitive volume, and the ratio between the energy deposited at the peak location and at the entrance of the sensitive volume are approximately the same for all the physics configurations and Geant4 versions, as illustrated in Fig. 22 and 23. The 95% confidence interval for the mean value deriving from Geant4 9.3 with *G4hMultipleScattering* is shown in the figures to appreciate quantitatively the spread of the results. These observations

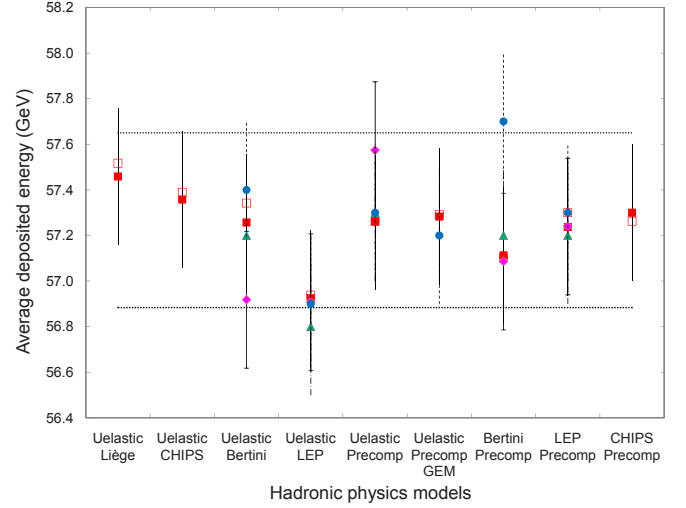


Fig. 22. Average energy deposition per proton entering the sensitive volume calculated with various Geant4 versions and physics configurations: Geant4 9.3 (red squares), Geant4 9.2p03 (pink diamonds), Geant4 9.1 (blue circles) and Geant4 8.1p02 (green triangles); the filled symbols correspond to simulations activating the *G4MultipleScattering* multiple scattering process, while the empty ones correspond to the activation of *G4hMultipleScattering*. The upper and lower lines of the horizontal axis identify respectively the hadronic elastic and inelastic scattering model in each simulation configuration; the other physics options, apart from the multiple scattering under test, were as in Table II. The dotted lines represent the 95% confidence interval for the mean value of the various Geant4 9.3 physics configurations associated with *G4hMultipleScattering*. The average energy deposition derives from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water.

suggest that the detailed features of the energy deposition in the water volume are insensitive to the physics options selected in the simulation, including multiple scattering, and to the evolutions of Geant4 software.

The acceptance, i.e. the fraction of protons reaching the sensitive volume, out of all the primary generated ones, is plotted in Fig. 24 for different physics configurations and Geant4 versions. Various sources can affect it: inelastic nuclear reactions, which remove protons from the beam prior to reaching the sensitive volume, and nuclear elastic and multiple Coulomb scattering, which modify the protons' direction along with their passage through matter.

The acceptance appears roughly constant in Fig. 24 for the various hadronic models, within the set of simulations associated with a given multiple scattering option and Geant4 version; therefore, the features of these models can be excluded as a source of significant differences. Complementary tests, whose results are not reported in Fig. 24, show that the acceptance is not significantly sensitive to alternative stopping power models either. The multiple scattering algorithm appears the most probable source of the observed differences.

Fig. 21 and 24 suggest a correlation between the total energy deposited in the sensitive volume and the acceptance. This effect was evaluated by means of Pearson's correlation coefficient [177]; the null hypothesis consists of assuming no correlation between these quantities. The correlation coefficient, calculated over all physics configurations and Geant4 versions examined in this study, is 0.965; the null hypothesis is rejected with 0.0001 significance. On the other hand, no

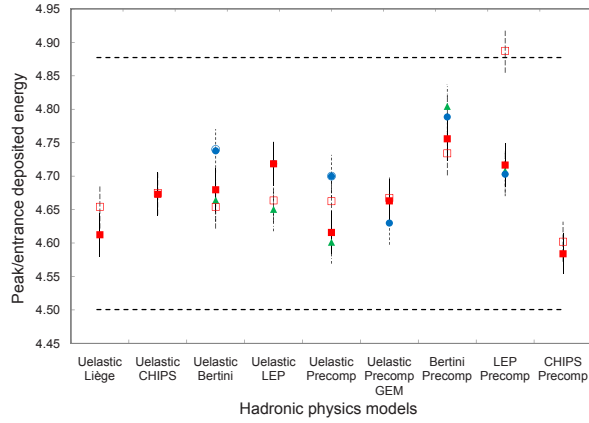


Fig. 23. Ratio of the energy deposition at the Bragg peak location and at the entrance of the sensitive volume, deriving from various Geant4 versions and physics configurations: Geant4 9.3 (red squares), Geant4 9.2p03 (pink diamonds), Geant4 9.1 (blue circles) and Geant4 8.1p02 (green triangles); the filled symbols correspond to simulations activating the *G4MultipleScattering* multiple scattering process, while the empty ones correspond to the activation of *G4hMultipleScattering*. The upper and lower lines of the horizontal axis identify respectively the hadronic elastic and inelastic scattering model in each simulation configuration; the other physics options, apart from the multiple scattering under test, were as in Table II. The dashed lines represent the 95% confidence interval for the mean value of the various Geant4 9.3 physics configurations associated with *G4hMultipleScattering*. The energy deposition derives from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water.

correlation of the total energy deposition is observed with the average energy deposited per proton, nor with the peak over entrance ratio: the corresponding correlation coefficients are 0.120 and 0.151; these values lead to not rejecting the null hypothesis with 0.1 significance.

These results hint that the observed discrepancies in the longitudinal energy deposit distributions are related to effects due to multiple scattering in the beam line, rather than to physics modeling effects in the water volume. Geant4 multiple scattering implementation encompasses various empirical parameters [74], whose settings are characterized by epistemic uncertainties; presumably, the observed effects are associated with different angular distribution (including backscattering) and lateral displacement of the scattered particle implemented in the various Geant4 multiple scattering options and versions, and the variations of empirical parameters governing the algorithm.

This finding stresses the importance of accurately modeling the beam line geometry and material composition for accurate calculation of the energy deposited in the sensitive volume. It also highlights the importance of correctly simulating particle interactions not only in the sensitive parts of the experimental set-up, but also in its passive components, since the latter appear to be responsible for significant systematic effects on the energy deposited in the sensitive volume.

In hadron therapy practice, proton depth dose profiles are usually normalized to a reference value (at the peak or at the entrance of the sensitive volume), or to the integral of

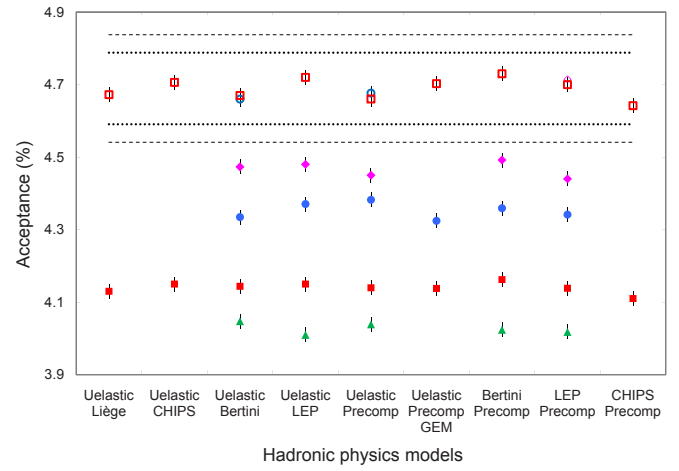


Fig. 24. Percentage of primary protons (acceptance) reaching the sensitive volume deriving from various Geant4 versions and physics configurations: Geant4 9.3 (red squares), Geant4 9.2p03 (pink diamonds), Geant4 9.1 (blue circles) and Geant4 8.1p02 (green triangles); the filled symbols correspond to simulations activating the *G4MultipleScattering* multiple scattering process, while the empty ones correspond to the activation of *G4hMultipleScattering*. The upper and lower lines of the horizontal axis identify respectively the hadronic elastic and inelastic scattering model in each simulation configuration; the other physics options, apart from the multiple scattering under test, were as in Table II. The dashed and dotted lines represent respectively the 99.9% and 99% confidence intervals for the mean value of the various Geant4 9.3 physics configurations of associated with *G4hMultipleScattering*. The simulation involves from one million primary protons with $\langle E \rangle = 63.95$ MeV and $\sigma = 300$ keV incident on water.

the dose, as discussed in [44]. The same goodness-of-fit tests reported in Table VI were performed after normalizing the energy deposition profiles to the total energy collected in the sensitive volume: they failed to reject the null hypothesis of compatibility with 0.1 significance, which was rejected in the comparison of the original (non normalized) distributions. This analysis demonstrates that normalized distributions are insensitive to the large differences exhibited by the various models on an absolute scale. Therefore, comparisons like the one in Fig. 7 of [67], concerning an experimental Bragg peak and one simulated with Geant4 9.0, both normalized to 1, are of limited usefulness to clarify the issues that emerged in the previous analysis.

Further tests were performed, activating the specific *G4eMultipleScattering* process for electron multiple scattering, released in Geant4 9.3; no effects were observed on the longitudinal energy deposition.

The authors found only limited documentation in the literature concerning the experimental validation of proton multiple scattering implementations in recent Geant4 versions; the comparisons with experimental data reported in [178] concern muons and electrons and are not pertinent to the use case object of this investigation, which concerns protons. Therefore, this process is a source of epistemic uncertainty in the simulation; the analysis described in this paper shows that this uncertainty could determine large systematic effects in critical use cases. Further experimental measurements would be useful for the validation of Geant4 multiple scattering in use cases similar to the one considered in this paper; in particular, experimental data suitable to clarify the interplay

between energy deposition measurements and features of the multiple scattering algorithm would be beneficial. Ideally, the experiment should be able to measure effects related to backscattering and lateral displacement, which could be responsible for the discrepancies in the proton acceptance observed with the various algorithm implementations examined in this paper.

VI. COMPUTATIONAL PERFORMANCE

The extensive survey of physics models and parameters relevant to the problem domain documented in the previous sections provides guidance for Geant4-based simulations concerning similar use cases. The computational performance of the available physics options is a relevant parameter in simulation applications, especially considering that some previous analyses demonstrate the equivalence of some of them regarding the physical features they produce. Therefore the analysis is complemented here by some information on the associated computational performance in the use case described in this paper, which can be useful to experimentalists in their Geant4-based applications.

The results reported in Table VII, related to Geant4 9.3 version, show the average simulation time per primary generated event in each physics configuration; they derive from the productions for the analysis described in the previous sections. The content of Table VII should not be considered as measurements of Geant4 computational performance in absolute terms: the application code contained analysis features, such as filling a large number of histograms, which added an additional burden to the execution with respect to the time strictly needed for particle transport; moreover no effort was invested in the optimization of the application code. However, since all the simulations reported in that table were run on identical hardware and platforms, the measured execution times are interesting for relative comparisons of the computational performance of the various physics configurations in the use case object of this study.

The results reported in Table VII involve the *ICRU49* proton stopping power model; simulations involving the *Ziegler77*, *Ziegler85* and *Ziegler2000* models are slightly slower. Simulations involving *G4hMultipleScattering* require approximately 5% more CPU time than those involving *G4MultipleScattering*; however, the larger acceptance associated with this multiple scattering model requires longer computations to track a greater number of particles in the sensitive volume.

It is worth remarking that accounting for nuclear interactions in the simulation application described in this paper increases the computational time consumption by approximately 57%, with respect to considering electromagnetic interactions only.

Based on Table VII, one can observe that the hadronic elastic scattering models exhibit similar computational performance, with the exception of the CHIPS model, which is significantly slower; among the hadronic inelastic models, the Liège cascade and the LEP ones are faster than the other options.

TABLE VII
AVERAGE CPU TIME PER PRIMARY GENERATED EVENT IN VARIOUS
PHYSICS CONFIGURATIONS

| Hadronic elastic | Hadronic inelastic | CPU time (ms) |
|------------------|----------------------------|-----------------|
| Bertini-elastic | Precompound | 254.0 \pm 0.3 |
| LEP | Precompound | 255.1 \pm 0.3 |
| CHIPS-elastic | Precompound | 293.3 \pm 0.3 |
| U-elastic | Precompound | 254.3 \pm 0.3 |
| U-elastic | Precompound-GEM | 251.1 \pm 0.3 |
| U-elastic | Precompound-Fermi break-up | 255.8 \pm 0.3 |
| U-elastic | Binary cascade | 261.3 \pm 0.3 |
| U-elastic | Bertini cascade | 251.7 \pm 0.3 |
| U-elastic | Liège cascade | 223.1 \pm 0.2 |
| U-elastic | LEP | 225.4 \pm 0.2 |
| U-elastic | CHIPS-inelastic | 250.4 \pm 0.3 |

VII. CONCLUSION

A number of epistemic uncertainties have been identified in a survey of Geant4 physics models pertinent to the simulation of proton depth dose, which broadly represent the variety of approaches to describe proton interactions with matter in the energy range up to approximately 100 MeV.

In the electromagnetic domain, the epistemic uncertainties affecting the value of the water mean ionization potential and proton stopping powers derive from lack of consensus among various theoretical and experimental references documented in the literature; they generate significant systematic effects on the longitudinal pattern of energy deposit in the sensitive volume, namely on the depth of the Bragg peak.

The epistemic uncertainties affecting the hadronic components of the simulation are related to the intrinsic differences of the modeling approaches and empirical parameters they contain; the limited validation of the models, and the unclear distinction between the processes of calibration and validation in the few published comparisons with experimental data, are the main sources of such uncertainties. Their effects on the longitudinal energy deposit are comparable with experimental uncertainties typical of proton therapy; the largest differences concern secondary particle spectra. A significant effect was observed in relation to the mode of nuclear deexcitation; in this respect, there is a consensus towards modeling it through Fermi break-up for light nuclei and evaporation for heavier ones. This approach is implemented in some Monte Carlo codes (e.g. MCNP and FLUKA), while it is not adopted by default in Geant4; users of this code would benefit from implementing appropriate settings in their Geant4-based applications to activate Fermi break-up for the deexcitation of light nuclei, if their simulation use cases are prone to be affected by the systematic effects highlighted in this study.

The analysis shows how the sensitivity of the simulation to epistemic uncertainties cannot be determined in absolute terms, rather it depends on the experimental application environment. The relatively large differences in the Bragg peak profile associated with the set of electromagnetic options are practically irrelevant in clinical practice, which tolerates adjustments of the beam parameters to reproduce a reference proton range. However, these differences are relevant to applications where a predictive role is expected from the simulation, such as Monte Carlo based treatment planning

TABLE VI

P-VALUE OF GOODNESS-OF-FIT TESTS COMPARING LONGITUDINAL ENERGY DEPOSITION PROFILES DERIVING FROM VARIOUS GEANT4 VERSIONS AND PHYSICS CONFIGURATIONS

| Compared software versions | Physics configuration | Kolmogorov Smirnov | Anderson Darling | Cramer von Mises |
|---|-----------------------|--------------------|------------------|------------------|
| 9.3 (G4hMultipleScattering) 9.3 (G4MultipleScattering) | Uelastic Bertini | < 0.001 | < 0.001 | < 0.001 |
| | Uelastic LEP | < 0.001 | < 0.001 | < 0.001 |
| | Uelastic Precompound | < 0.001 | < 0.001 | < 0.001 |
| | Bertini Precompound | < 0.001 | < 0.001 | < 0.001 |
| | LEP Precompound | < 0.001 | < 0.001 | < 0.001 |
| 9.3 (G4hMultipleScattering) 9.2 (G4MultipleScattering) | Uelastic Bertini | 0.219 | 0.046 | 0.110 |
| | Uelastic LEP | 0.074 | 0.013 | 0.047 |
| | Uelastic Precompound | 0.170 | 0.037 | 0.111 |
| | Bertini Precompound | 0.054 | 0.012 | 0.044 |
| | LEP Precompound | 0.098 | 0.019 | 0.057 |
| 9.3 (G4hMultipleScattering) 9.1 (G4MultipleScattering) | Uelastic Bertini | 0.009 | 0.001 | 0.004 |
| | Uelastic LEP | 0.002 | < 0.001 | 0.002 |
| | Uelastic Precompound | 0.014 | 0.002 | 0.008 |
| | Bertini Precompound | 0.004 | < 0.001 | 0.002 |
| | LEP Precompound | 0.006 | 0.001 | 0.006 |
| 9.3 (G4hMultipleScattering) 8.1 (G4MultipleScattering) | Uelastic Bertini | < 0.001 | < 0.001 | < 0.001 |
| | Uelastic LEP | < 0.001 | < 0.001 | < 0.001 |
| | Uelastic Precompound | < 0.001 | < 0.001 | < 0.001 |
| | Bertini Precompound | < 0.001 | < 0.001 | < 0.001 |
| | LEP Precompound | < 0.001 | < 0.001 | < 0.001 |
| 9.3 (G4MultipleScattering) 9.2 (G4MultipleScattering) | Uelastic Bertini | 0.006 | < 0.001 | 0.004 |
| | Uelastic LEP | 0.001 | < 0.001 | 0.002 |
| | Uelastic Precompound | 0.006 | < 0.001 | 0.003 |
| | Bertini Precompound | < 0.001 | < 0.001 | 0.001 |
| | LEP Precompound | 0.009 | < 0.001 | 0.005 |
| 9.3 (G4MultipleScattering) 9.1 (G4MultipleScattering) | Uelastic Bertini | 0.277 | 0.051 | 0.113 |
| | Uelastic LEP | 0.039 | 0.011 | 0.043 |
| | Uelastic Precompound | 0.054 | 0.012 | 0.040 |
| | Bertini Precompound | 0.039 | 0.007 | 0.028 |
| | LEP Precompound | 0.130 | 0.030 | 0.071 |
| 9.3 (G4MultipleScattering) 8.1 (G4MultipleScattering) | Uelastic Bertini | 0.803 | 0.505 | 0.722 |
| | Uelastic LEP | 0.277 | 0.119 | 0.270 |
| | Uelastic Precompound | 0.515 | 0.232 | 0.475 |
| | Bertini Precompound | 0.219 | 0.072 | 0.179 |
| | LEP Precompound | 0.426 | 0.150 | 0.297 |
| 9.2 (G4MultipleScattering) 9.1 (G4MultipleScattering) | Uelastic Bertini | 0.426 | 0.135 | 0.286 |
| | Uelastic LEP | 0.709 | 0.281 | 0.395 |
| | Uelastic Precompound | 0.884 | 0.418 | 0.548 |
| | Bertini Precompound | 0.709 | 0.324 | 0.478 |
| | LEP Precompound | 0.426 | 0.269 | 0.516 |
| 9.2 (G4hMultipleScattering) 8.1 (G4MultipleScattering) | Uelastic Bertini | < 0.001 | < 0.001 | < 0.001 |
| | Uelastic LEP | < 0.001 | < 0.001 | < 0.001 |
| | Uelastic Precompound | < 0.001 | < 0.001 | < 0.001 |
| | Bertini Precompound | < 0.001 | < 0.001 | < 0.001 |
| | LEP Precompound | < 0.001 | < 0.001 | < 0.001 |
| 9.1 (G4MultipleScattering) 8.1 (G4MultipleScattering) | Uelastic Bertini | 0.020 | 0.004 | 0.016 |
| | Uelastic LEP | 0.001 | < 0.001 | 0.001 |
| | Uelastic Precompound | 0.003 | < 0.001 | 0.003 |
| | Bertini Precompound | < 0.001 | < 0.001 | < 0.001 |
| | LEP Precompound | 0.006 | < 0.001 | 0.003 |

systems, currently the object of active research, or radiation protection. The different secondary particle spectra deriving from the range of available hadronic options do not affect the main parameter of clinical interest, i.e. the depth dose distribution, but they are relevant to other aspects of radiation exposure.

By far the largest effects of physics-related epistemic uncertainties in the simulation of proton depth dose are observed in relation to modeling multiple scattering in the beam line. However, even these effects are relevant only to use cases where the simulation is invested with predictive role regarding the absolute dose released to the target; otherwise, common practices, like the normalization of the simulated dose to a ref-

erence value, would mask the epistemic uncertainty associated with the empirical parameters used to model this process.

The analysis also highlights the importance of a knowledge of the whole simulation system regarding the effects visible in the sensitive volume. Interactions in the beam line affect the spectrum of the protons reaching the sensitive volume and the dose released to it; lack of knowledge of construction details of the beam line, or epistemic uncertainties in modeling particle interactions in the passive components of the system, are prone to bias the simulation outcome.

The results documented in this paper about the different observables produced by Geant4 physics options identify some experimental requirements for the discrimination of their

features and their validation. Experimental measurements of adequate accuracy could reduce the epistemic uncertainties evidenced in the electromagnetic domain; relevant data could derive either from a thorough survey of the existing literature, or from new, dedicated measurements. In this respect, it is worthwhile to recall the valuable reference role for the validation of electron simulation played by the high precision measurements of [179] and [180], which were originally motivated by the validation of the ITS (Integrated Tiger Series) [181] Monte Carlo code; similar measurements concerning protons would be useful to reduce epistemic uncertainties.

The sensitivity analysis documented in the previous sections also provides guidance to design meaningful test cases for inclusion in the test process of Monte Carlo systems. The identification of distributions which expose distinctive features of the physics models, as well as of others, which are prone to hide them, is especially useful to designing test cases relevant to monitoring the effects of changes in some critical parts of the code.

The analysis presented in this paper is a first attempt at estimating quantitatively the impact of epistemic uncertainties on the considered use case; further refinements would contribute to better understanding the problem. So far, the analysis has considered each source of epistemic uncertainty individually; nevertheless, it would be worthwhile to evaluate their combinations, since several systematic contributions could accumulate their effects to bias the final simulation result. More refined treatments, e.g. based on the theory of evidence, could shed additional light on the problem; these methods would be especially useful if practical constraints hinder the availability of further experimental measurements to reduce the current uncertainties.

The identification of the epistemic uncertainties embedded in a large-scale simulation code is far from trivial; design methods facilitating their identification at early stages of the software development, and their management in sensitivity analyses, would be beneficial. To the best of the authors' knowledge, this issue has not been studied yet in the context of Monte Carlo simulation; techniques like aspect oriented programming could provide useful paradigms to address it, and the inclusion of epistemic uncertainties in the traceability process, in the context of a rigorous software process discipline, would effectively support their handling in complex software systems.

Although this paper illustrates the problem of epistemic uncertainties in a specific simulation use case, the issue it addresses goes beyond the limited application domain considered in this initial study. Regarding the simulation of low energy proton interactions, the epistemic uncertainties discussed in this paper and their effects are likely to affect other experimental domains as well: from the exposure of electronic components and astronauts to the space radiation environment, to the problem of radiation monitoring at particle accelerators.

More generally, the issue of identifying and quantifying epistemic uncertainties, and their contribution to the overall reliability of simulation systems, permeates all Monte Carlo application domains. Monte Carlo simulation - not only for

particle transport in detectors, but also for event generators - is expected to play a critical role in the physics analysis of LHC data, which involves energies not yet covered by any experimental measurements in controlled laboratory environments; the development of sound methods and tools to deal with the epistemic uncertainties embedded in LHC simulation software appears a major task for the coming years in support of LHC physics results.

ACKNOWLEDGMENT

The authors thank Andreas Pfeiffer for his significant help with data analysis tools throughout the project; Katsuya Amako, Sergio Bertolucci, Luciano Catani, Gloria Corti, Andrea Dotti, Gunter Folger, Simone Giani, Vladimir Grichine, Aatos Heikkinen, Alexander Howard, Vladimir Ivanchenko, Mikhail Kossov, Vicente Lara, Katia Parodi, Alberto Ribon, Takashi Sasaki, Vladimir Uzhinsky and Hans-Peter Wellisch for valuable discussions, and Anita Hollier for proofreading the manuscript.

CERN Library's support has been essential to this study; the authors are especially grateful to Tullio Basaglia.

INFN Genova Computing Service (Alessandro Brunengo, Mirko Corsu, Paolo Lantero and Francesco Saffioti) provided helpful technical assistance with the simulation production.

The authors do not intend to express criticism, nor praise regarding any of the Monte Carlo codes mentioned in this paper; the purpose of the paper is limited to documenting technical results.

REFERENCES

- [1] W. P. Levin, H. Kooy, J. S. Loeffler, and T. F. DeLaney, "Proton beam therapy", *Brit. J. Cancer*, vol. 93, pp. 849-854, 2005.
- [2] X-5 Monte Carlo Team, "MCNP - A General Monte Carlo N-Particle Transport Code, Version 5", Los Alamos National Laboratory Report LA-UR-03-1987, 2003, revised 2005.
- [3] R. A. Forster et al., "MCNP Version 5", *Nucl. Instrum. Meth. B*, vol. 213, pp. 82-86, 2004.
- [4] J. S. Hendricks et al., "MCNPX, Version 26c", Los Alamos National Laboratory Report LA-UR-06-7991, 2006.
- [5] "GEANT Detector Description and Simulation Tool", CERN Program Library Long Writeup W5013, 1995.
- [6] S. Agostinelli et al., "Geant4 - a simulation toolkit" *Nucl. Instrum. Meth. A*, vol. 506, no. 3, pp. 250-303, 2003.
- [7] J. Allison et al., "Geant4 Developments and Applications" *IEEE Trans. Nucl. Sci.*, vol. 53, no. 1, pp. 270-278, 2006.
- [8] A. V. Dementyev and N. M. Sobolevsky, "SHIELD universal Monte Carlo hadron transport code: scope and applications", *Radiat. Meas.*, vol. 30, no. 5, pp. 553-557, 1999.
- [9] I. Gudowska, P. Andreo, and N. M. Sobolevsky, "Secondary Particle Production in Tissue-like and Shielding Materials for Light and Heavy Ions Calculated with the Monte-Carlo Code SHIELD-HIT", *J. Radiat. Res.*, vol. 43 Suppl., pp. 93-97, 2002.
- [10] G. Battistoni et al., "The FLUKA code: description and benchmarking", *AIP Conf. Proc.*, vol. 896, pp. 31-49, 2007.
- [11] A. Ferrari et al., "Fluka: a multi-particle transport code", Report CERN-2005-010, INFN/TC-05/11, SLAC-R-773, Geneva, Oct. 2005.
- [12] H. Iwase, K. Niita, and T. Nakamura, "Development of general-purpose particle and heavy ion transport Monte Carlo code", *J. Nucl. Sci. Technol.*, vol. 39, no. 11, pp. 1142-1151, 2002.
- [13] W. L. Oberkampf, S. M. DeLand, B. M. Rutherford, K. V. Diegert, and K. F. Alvin, "Error and uncertainty in modeling and simulation", *Reliab. Eng. Syst. Safety*, vol. 75, no. 3, pp. 333-357, 2002.
- [14] W. L. Oberkampf, T. G. Trucano, and C. Hirsch, "Verification, validation, and predictive capability in computational engineering and physics", *Appl. Mech. Rev.*, vol. 57, no. 5, pp. 345-384, 2004.

- [15] J. Hérault, N. Iborra, B. Serrano, and P. Chauvel, "Monte Carlo simulation of a protontherapy platform devoted to ocular melanoma", *Med. Phys.*, vol. 32, pp. 910-919 2005.
- [16] D. E. Bonnett, A. Kacperek, M. A. Sheen, R. Goodall, and T. E. Saxton, "The 62 MeV proton beam for the treatment of ocular melanoma at Clatterbridge", *Brit. J. Radiol.*, vol. 66, pp. 907-914, 1993.
- [17] S. Höcht et al., "Proton Therapy of Uveal Melanomas in Berlin", *Strahl. Onk.*, vol. 180, no. 7, pp. 419-424, 2004.
- [18] G. A. P. Cirrone et al., "A 62 MeV proton beam for the treatment of ocular melanoma at Laboratori Nazionali del Sud-INFN (CATANIA)", in *IEEE Trans. Nucl. Sci.*, vol. 51, no. 3, pp. 860-865, 2004.
- [19] T. G. Trucano, L. P. Swiler, T. Igusa, W. L. Oberkampf, and M. Pilch, "Calibration, validation, and sensitivity analysis: Whats what", *Reliab. Eng. Syst. Safety*, vol. 91, no. 10-11, pp. 1331-1357, 2006.
- [20] J. C. Helton, "Alternative representations of epistemic uncertainty", *Reliab. Eng. Syst. Safety*, vol. 84, no. 1-3, pp. 1-10, 2004.
- [21] G. Shafer, "A Mathematical Theory of Evidence", Princeton University Press, 1976.
- [22] A. Saltelli, K. Chan, and M. Scott M, editors, "Mathematical and statistical method for sensitivity analysis", New York: Wiley; 2001.
- [23] K. Diegert, S. Klenke, G. Novotny, R. Paulsen, M. Pilch, and T. Trucano, "Toward a More Rigorous Application of Margins and Uncertainties within the Nuclear Weapons Life Cycle - A Sandia Perspective", Sandia National Laboratories Report No. SAND2007-6219, Albuquerque, 2007.
- [24] A. Lechner and M. G. Pia, "Analysis of Geant4 simulations of proton depth dose profiles for radiotherapy applications", in *IEEE Nucl. Sci. Symp. Conf. Rec.* 2008, pp. 875-882, 2008.
- [25] H. Paganetti and B. Gottschalk, "Test of GEANT3 and GEANT4 nuclear models for 160 MeV protons stopping in CH₂", *Med. Phys.*, vol. 30, no. 7, pp. 1926-1931, 2003.
- [26] M. Fippel and M. Soukup, "A Monte Carlo dose calculation algorithm for proton therapy" *Med. Phys.*, vol. 31, no. 8, pp. 2263-2273, 2004.
- [27] H. Paganetti, H. Jiang, S.-Y. Lee, and H. M. Kooy, "Accurate Monte Carlo simulations for nozzle design, commissioning and quality assurance for a proton radiation therapy facility" *Med. Phys.*, vol. 31, no. 7, pp. 2107-2118, 2004.
- [28] A. B. Rosenfeld, A. J. Wroe, I. M. Cornelius, M. Reinhard, and D. Alexiev, "Analysis of inelastic interactions for therapeutic proton beams using Monte Carlo simulation", *IEEE Trans. Nucl. Sci.*, vol. 51, no. 6, pp. 3019-3025, 2004.
- [29] A. J. Wroe, I. M. Cornelius, and A. B. Rosenfeld, "The role of nonelastic reactions in absorbed dose distributions from therapeutic proton beams in different medium", *Med. Phys.*, vol. 32, no. 1, pp. 37-41, 2005.
- [30] G. A. P. Cirrone et al., "Implementation of a New Monte Carlo GEANT4 Simulation Tool for the Development of a Proton Therapy Beam Line and Verification of the Related Dose Distributions", *IEEE Trans. Nucl. Sci.*, vol. 52, no. 1, pp. 262-265, 2005.
- [31] T. Aso et al., "Verification of the Dose Distributions With GEANT4 Simulation for Proton Therapy", *IEEE Trans. Nucl. Sci.*, vol. 52, no. 4, pp. 896-901, 2005.
- [32] A. Kim, J. W. Kim, I. Hahn, N. Schreuder, and J. Farr, "Simulations of Therapeutic Proton Beam Formation with Beam GEANT4" *J. Kor. Phys. Soc.*, vol. 47, no. 2, pp. 197-201, Aug. 2005.
- [33] C. Baker, D. Shipley, H. Palmans, and A. Kacperek, "Monte carlo modelling of a clinical proton beam-line for the treatment of ocular tumours", *Nucl. Instrum. Meth. A*, vol. 562, pp. 1005-1008, 2006.
- [34] G. A. P. Cirrone et al., "Validation of Geant4 Physics Models for the Simulation of the Proton Bragg Peak", in *IEEE 2006 Nucl. Sci. Symp. Conf. Rec.*, pp. 788-792.
- [35] C. Zacharatou Jarlskog and H. Paganetti, "Physics settings for using the Geant4 Toolkit in proton therapy", *IEEE Trans. Nucl. Sci.*, vol. 55, no. 3, pp. 1018-1025, Jun. 2008.
- [36] K. Amako et al., "Comparison of Geant4 electromagnetic physics models against the NIST reference data", *IEEE Trans. Nucl. Sci.*, vol. 52, no. 4, pp. 910-918, 2005.
- [37] M. P. W. Chin and N. M. Spyrou, Non-convergence of Geant4 hadronic models for 10 and 30 MeV protons in 18O and 14N Applied Radiation and Isotopes Volume 67, Issue 3, March 2009, Pages 406-414 Proceedings of the first international conference on biomedical applications of high energy ion beams
- [38] J. M. Quesada, M. A. Cortes, A. Howard, G. Folger, and V. N. Ivanchenko, "Improvements of preequilibrium and evaporation models in Geant4", in *IEEE Nucl. Sci. Symp. Conf. Rec.*, pp. 847-849, 2008.
- [39] V. N. Ivanchenko and A. Ivantchenko, "Testing suite for validation of Geant4 hadronic generators", *J. Phys.: Conf. Ser.*, vol. 119, pp. 032026, 2008.
- [40] J. Apostolakis et al., "Progress in hadronic physics modelling in Geant4", *J. Phys.: Conf. Ser.*, vol. 160, pp. 012073, 2009.
- [41] W. Newhauser, N. Koch, S. Hummel, M. Ziegler, and U. Titt, "Monte Carlo simulations of a nozzle for the treatment of ocular tumours with high-energy proton beams", *Phys. Med. Biol.*, vol. 50, pp. 5229-5249, 2005.
- [42] I. Gudowska, N. Sobolevsky, P. Andreo, D. Belkic and A. Brahme, "Ion beam transport in tissue-like media using the Monte Carlo code SHIELD-HIT", *Phys. Med. Biol.*, vol. 49, pp. 1933-1958, 2004.
- [43] F. D. Brooks, D. T. L. Jones, C. C. Bowley, J. E. Symons, A. Buffler, and M. S. Allie, "Energy spectra in the NAC proton therapy beam", *Radiat. Prot. Dosim.*, vol. 70, pp. 477-480, 1997.
- [44] A. Lechner and M. G. Pia, "Effect of normalization algorithms on the analysis of Bragg peak profiles", *IEEE Trans. Nucl. Sci.*, vol. 55, no. 6, pp. 3544-3549, 2008.
- [45] C. Agodi et al., "The INFN TPS project", *Nuovo Cimento C*, vol. 31, no. 1, pp. 99-108, 2008.
- [46] N. Reynaert et al., "Monte Carlo treatment planning for photon and electron beams", *Radiat. Phys. Chem.*, vol. 76, no. 4, pp. 643-686, 2007.
- [47] N. Tilly, "Monte Carlo calculations for proton therapy planning, is it worth the effort?", *Radiotherapy Oncol.*, vol. 76, Suppl. 2, pp. S17-S18, 2005.
- [48] S. Webb, "The contribution, history, impact and future of physics in medicine", *Acta Oncol.*, vol. 48, no. 2, pp. 169-177, 2009.
- [49] S. Chauvie et al., "Geant4 electromagnetic physics", in *Proc. Monte Carlo Conference, Lisbon*, Nov. 2000.
- [50] S. Chauvie, G. Depaola, V. Ivanchenko, F. Longo, P. Nieminen and M. G. Pia, "Geant4 Low Energy Electromagnetic Physics", in *Proc. Computing in High Energy and Nuclear Physics*, Beijing, China, pp. 337-340, 2001.
- [51] S. Chauvie et al., "Geant4 Low Energy Electromagnetic Physics", in *Conf. Rec. 2004 IEEE Nucl. Sci. Symp.*, N33-165.
- [52] S. Giani, V. N. Ivanchenko, G. Mancinelli, P. Nieminen, M. G. Pia, and L. Urban, "Geant4 simulation of energy losses of slow hadrons", *INFN/AE-99/20*, Frascati, 1999.
- [53] J. Lindhard and M. Scharff, "Energy loss in matter by fast particles of low charge", *Mat. Fys. Medd. Dan. Vid. Selsk.*, vol. 27, no. 15, pp. 1, 1953.
- [54] M. J. Berger et al., "Stopping Powers and Ranges for Protons and Alpha Particles", *ICRU Report 49*, Bethesda, 1993.
- [55] H. H. Andersen and J. F. Ziegler, "The Stopping and Ranges of Ions in Matter", Ed. Oxford: Pergamon Press, 1977.
- [56] J. F. Ziegler, J. P. Biersack, and U. Littmark, "The Stopping and Ranges of Ions in Solids", Ed. Oxford: Pergamon Press, 1985.
- [57] J. F. Ziegler, "SRIM-2000", [Online]. Available: <http://www.research.ibm.com/ionbeams/SRIM/SRIMLEGL.HTM>.
- [58] J. Apostolakis, S. Giani, M. Maire, P. Nieminen, M.G. Pia, L. Urban, "Geant4 low energy electromagnetic models for electrons and photons" *INFN/AE-99/18*, Frascati, 1999.
- [59] S. Giani, V. N. Ivanchenko, G. Mancinelli, P. Nieminen, M. G. Pia and L. Urban, "Geant4 simulation of energy losses of ions", *INFN/AE-99/21*, Frascati, 1999.
- [60] S. T. Perkins et al., "Tables and Graphs of Electron-Interaction Cross Sections from 10 eV to 100 GeV Derived from the LLNL Evaluated Electron Data Library (EEDL)", UCRL-50400 Vol. 31, 1997.
- [61] D. Cullen et al., "EPDL97, the Evaluated Photon Data Library", UCRL-50400, Vol. 6, Rev. 5, 1997.
- [62] J. Baro, J. Sempau, J. M. Fernández-Varea, and F. Salvat, "PENELOPE, an algorithm for Monte Carlo simulation of the penetration and energy loss of electrons and positrons in matter", *Nucl. Instrum. Meth. B*, vol. 100, no. 1, pp. 31-46, 1995.
- [63] S. Guatelli, A. Mantero, B. Mascialino, P. Nieminen, and M. G. Pia, "Geant4 Atomic Relaxation", *IEEE Trans. Nucl. Sci.*, vol. 54, no. 3, pp. 585-593, 2007.
- [64] M. J. Berger et al., "Stopping Powers and Ranges for Protons and Alpha Particles", *ICRU Report 49*, Bethesda, 1993.
- [65] H. Burkhardt et al., "Geant4 Standard Electromagnetic Package", in *Proc. 2005 Conf. on Monte Carlo Method: Versatility Unbounded in a Dynamic Computing World*, Am. Nucl. Soc., USA, 2005.
- [66] K. Lassila-Perini and L. Urban, "Geant4 Standard Electromagnetic Package", *Nucl. Instrum. Meth. A*, vol. 362, no. 2-3, pp. 416-422, 1995.

- [67] J. Apostolakis et al., "The performance of the Geant4 Standard EM package for LHC and other applications", *J. Phys. Conf. Series*, vol. 119, pp. 1-10, 2008.
- [68] J. Apostolakis et al., "Geometry and physics of the Geant4 toolkit for high and medium energy applications", *Radiat. Phys. Chem.*, vol. 78, no. 10, pp. 859-873, 2009.
- [69] A. Lechner, M.G. Pia, M. Sudhakar "Validation of Geant4 low energy electromagnetic processes against precision measurements of electron energy deposit", *IEEE Trans. Nucl. Sci.*, vol. 56, no. 2, pp. 398-416, 2009.
- [70] L. Urban, "Multiple scattering model in Geant4", CERN-OPEN-2002-070, Geneva, Switzerland, 2002.
- [71] L. Urban, "A model of multiple scattering in Geant4", CERN-OPEN-2006-077, Geneva, Switzerland, 2006.
- [72] H. W. Lewis, "Multiple Scattering in an Infinite Medium", *Phys. Rev.*, vol. 78, pp. 526-529, 1950.
- [73] Geant4 9.3 Release Notes, Online. Available: <http://cern.ch/geant4/support/ReleaseNotes4.9.3.html>
- [74] S. Elles, V. N. Ivanchenko, M. Maire and L. Urban, "Geant4 and Fano cavity: where are we?", *J. Phys. Conf. Series*, vol. 102, pp. 1-8, 2009.
- [75] R. Nartallo et al., "Radiation environment induced degradation on Chandra and implications for XMM", ESA Report, Esa/estec/tos-em/00-015/RN, Noordwijk, 2000.
- [76] W. J. Murray, "MuScat report - the muon scattering experiment", *J. Phys. G*, vol. 29, pp. 1595-1600, 2003.
- [77] P. Arce and M. Wadhwa, "Deviation in matter of 45 GeV muons in GEANT3 and GEANT4. A comparison with L3 data", CMS NOTE 2000/016, CERN, Geneva, 2000.
- [78] C. Rovelli, "Validation of the simulation of the CMS electromagnetic calorimeter using data", in *IEEE Nucl. Sci. Symp. Conf. Rec. 2008*, pp. 2792-2794, 2008.
- [79] K. Parodi and S. Squarcia, "Improvement of low-energy stopping power algorithms in the FLUKA simulation program", *Nucl. Instr. Meth. A*, vol. 456, no. 3, p. 352-368, 2001.
- [80] T. W. Armstrong and K. C. Chandler, "Stopping powers and ranges for muons, charged pions, protons, and heavy ions", *Nucl. Instr. Meth.*, vol. 113, no. 2, 15, pp. 313-314, 1973.
- [81] T. Goorley, R. E. Prael, and H. Grady, "Verification of Stopping Powers for Proton Transport in MCNP5", Los Alamos National Laboratory Report LA-UR-03-4446, 2003.
- [82] H. Nose, Y. Kase, N. Matsufuji, and T. Kanai, "Field size effect of radiation quality in carbon therapy using passive method", *Med. Phys.*, vol. 36, no. 3, pp. 870-875, 2009.
- [83] F. Bourhaleb et al., "A treatment planning code for inverse planning and 3D optimization in hadrontherapy", *Comp. Biol. Med.*, vol. 38, no. 9, pp. 990-999, 2008.
- [84] H. G. Hughes et al., "MCNP5 for proton radiography", *Radiat. Prot. Dos.*, vol. 116, no. 1-4, pp. 109-112, 2005.
- [85] K. Niita, T. Sato, H. Iwase, H. Nose, H. Nakashima, and L. Sihver, "PHITS - a particle and heavy ion transport code system", *Radiat. Meas.*, vol. 41, no. 9-10, pp. 1080-1090, 2006.
- [86] A. Ferrari, P. R. Sala, R. Guaraldi and F. Padoani, "An improved multiple scattering model for charged particle transport", *Nucl. Instr. Meth. B*, vol. 71, no. 4, pp. 412-426, 2002.
- [87] J. P. Wellisch, "Hadronic shower models in Geant4 - the frameworks", *Comp. Phys. Comm.*, vol. 140, no. 1, pp. 65-75, 2001.
- [88] V. Grichine, "Geant4 hadron elastic diffuse model", *Comp. Phys. Comm.*, vol. 181, pp. 921-927, 2010.
- [89] H. S. Fesefeldt, "The simulation of hadronic showers: physics and applications", PITHA-85-02, RWTH Aachen, Sep. 1985.
- [90] G. A. Lobov, N. V. Stepanov, A. A. Sibirtsev, and Yu. V. Trebukhovskii, "Statistical Simulation of Hadron and Light-Nuclei Interactions with Nuclei. Intranuclear Cascade Model", ITEP Preprint No. ITEP-91, Moscow, 1983.
- [91] D. H. Wright et al., "Recent Developments and Validations in Geant4 Hadronic Physics" in *AIP Conf. Proc.*, vol. 867, pp. 479-486, 2006.
- [92] M. Kossov, "Chiral-invariant phase space model", *Eur. Phys. J. A*, vol. 14, no. 3, pp. 265-269, 2002.
- [93] P. V. Degtyarenko, M. V. Kossov, and H.-P. Wellisch, "Chiral invariant phase space event generator" *Eur. Phys. J. A*, vol. 9, no. 3, pp. 411-420, 2003.
- [94] H. P. Wellisch and D. Axen, "Total reaction cross section calculations in proton-nucleus scattering", *Phys. Rev. C*, vol. 54, pp. 1329-1332, 1996.
- [95] V. S. Barashenkov and V. M. Maltsev "Cross sections for elementary particle interactions" *Fortschr. Phys.*, vol. 9, pp. 549-611, 1961.
- [96] S. Kox et al., "Trends of total reaction cross sections for heavy ion collisions in the intermediate energy range", *Phys. Rev. C*, vol. 35, pp. 1678-1691, 1987.
- [97] W. Q. Shen et al., "Total reaction cross section for heavy-ion collisions and its relation to the neutron excess degree of freedom", *Nucl. Phys. A*, vol. 491, no. 1, pp. 130-146, 1989.
- [98] R. K. Tripathi, F. A. Cucinotta, and J. W. Wilson, "Universal Parameterization of Absorption Cross Sections", NASA Technical Paper 3621, Langley Research Center, Hampton, Virginia, Jan. 1997.
- [99] R. K. Tripathi, F. A. Cucinotta, and J. W. Wilson, "Accurate universal parameterization of absorption cross sections", *Nucl. Instrum. Meth. B*, vol. 117, no. 4, pp. 347-349, 1996.
- [100] R. K. Tripathi, F. A. Cucinotta, and J.W. Wilson, " Universal Parameterization of Absorption Cross Sections - Light Systems", NASA Technical Paper 1999-209726, 1999.
- [101] R. K. Tripathi, F. A. Cucinotta, and J.W. Wilson, " Universal Parameterization of Absorption Cross Sections III - Light Systems", *Nucl. Instrum. Meth. B*, vol. 155, no. 4, pp. 349-356, 1999.
- [102] N. Amelin, "Physics and algorithms of the hadronic Monte-Carlo event generators: notes for a developer" Report CERN-IT-99-006, CERN, Geneva, 1999.
- [103] G. Folger, V.N. Ivanchenko, and J.P. Wellisch, "The Binary Cascade", *Eur. Phys. J. A*, vol. 21, pp. 407-417, 2004.
- [104] A. Heikkinen, N. Stepanov, Nikita, and J. P. Wellisch, "Bertini intra-nuclear cascade implementation in Geant4", in *Proc. of CHEP 2003 Int. Conf. Comp.in High En. and Nucl. Phys.*, La Jolla, 2003.
- [105] A. Heikkinen, "Implementing the Bertini intra-nuclear-cascade in the Geant4 hadronic framework", in *The Monte Carlo Method: Versatility Unbounded in a Dynamic Computing World*, on CD-ROM, Am. Nucl. Soc., 2005.
- [106] A. Boudard, J. Cugnon, S. Leray, and C. Volant, "Intranuclear cascade model for a comprehensive description of spallation reaction data", *Phys. Rev. C*, vol. 66, no. 4, pp. 044615, 2002.
- [107] V. Lara and J. P. Wellisch, "Pre-equilibrium and equilibrium decays in Geant4", in *Proc. CHEP 2000 Int. Conf. on Computing in High Energy and Nuclear Physics*, Padova, 2000.
- [108] J. J. Griffin, "Statistical Model of Intermediate Structure", *Phys. Rev. Lett.*, vol. 17, pp. 478-481, 1966.
- [109] J. J. Griffin, "Statistical Model of Intermediate Structure, erratum", *Phys. Lett. B*, vol. 19, pp. 57-57, 1967.
- [110] V. Weisskopf, "Statistics and Nuclear Reactions", *Phys. Rev.*, vol. 52, pp. 295 - 303, 1937.
- [111] V. F. Weisskopf and D. H. Ewing, "On the Yield of Nuclear Reactions with Heavy Elements", *Phys. Rev.*, vol. 57, pp. 472 - 485, 1940.
- [112] I. Dostrovsky, Z. Fraenkel, and G. Friedlander, "Monte Carlo Calculations of Nuclear Evaporation Processes. III. Applications to Low-Energy Reactions" *Phys. Review*, vol. 116, pp. 683-702, 1959.
- [113] S. Furihata, "Statistical analysis of light fragment production from medium energy proton-induced reactions", *Nucl. Instr. Meth. B*, vol. 171, no. 3, pp. 251-258, 2000.
- [114] V. D. Toneev and K. K. Gudima, "Particle emission in light and heavy ion reactions", *Nucl. Phys. A*, vol. 400, pp. 173-189, 1983.
- [115] K. K. Gudima, S. G. Mashnik, and V. D. Toneev, "Cascade-exciton model of nuclear reactions" *Nucl. Phys. A*, vol. 401, pp. 329-361, 1983.
- [116] H. W. Bertini, "Low-Energy Intranuclear Cascade Calculation", *Phys. Rev.*, vol. 131, pp. 1801-1821, 1963.
- [117] S. G. Mashnik and V. D. Toneev, "MODEX - the Program for Calculation of the Energy Spectra of Particles Emitted in the Reactions of Pre-Equilibrium and Equilibrium Statistical Decays", *Comm. JINR, P4-8417*, Dubna, 1974 .
- [118] S. G. Mashnik and A. J. Sierk, "Improved Cascade-Exciton Model of Nuclear Reactions", Los Alamos National Laboratory Report LA-UR-98-5999, 1998.
- [119] K. C. Chandler and T. W. Armstrong, "Operating instructions for the high-energy nucleon-meson transport code HETC", Oak Ridge National Laboratory Report ORNL-4744, 1972.
- [120] R. E. Prael and H. Lichtenstein, "User Guide to LCS: the LAHET Code System ", Los Alamos National Laboratory Report LA-UR-89-3014, 1989.
- [121] Y. Yariv and Z. Fraenkel, "Intranuclear cascade calculation of high-energy heavy-ion interactions", *Phys. Rev. C*, vol. 20, pp. 2227-2243, 1979.
- [122] Y. Yariv and Z. Fraenkel, "Intranuclear cascade calculation of high energy heavy ion collisions: Effect of interactions between cascade particles", *Phys. Rev. C*, vol. 24, pp. 488-494, 1981.

- [123] C. Zeitnitz and T. Gabriel, "The GEANT-CALOR interface and benchmark calculations of ZEUS test calorimeters", *Nucl. Instr. Meth. A*, vol. 349, no. 1, pp. 106-111, 1994.
- [124] J. J. Gaimard and K. H. Schmidt, "A reexamination of the abrasion-ablation model for the description of the nuclear fragmentation reaction", *Nucl. Phys. A*, vol. 531, no. 3-4, pp. 709-745, 2001.
- [125] A. Heikkinen, P. Kaitaniemi and A. Boudard, "Implementation of INCL cascade and ABLA evaporation codes in Geant4", *J. Phys.: Conf. Ser.*, vol. 119, pp. 032024, 2008.
- [126] J. Cugnon et al., "The INCL model for spallation reactions below 10 GeV", *Adv. Space Res.*, vol. 40, no. 9, pp. 1332-1338, 2007.
- [127] T. Aoust et al., "Recent extensions of the INCL4 model for spallation reactions", *Nucl. Instr. Meth. A*, vol. 562, no. 2, pp. 810-813, 2006.
- [128] S. Pedoux, J. Cugnon, A. Boudard, J. C. David and S. Leray, "Extension of INCL4 between 2 and 15 GeV", *Adv. Space Res.*, vol. 44, no. 8, pp. 926-933, 2009.
- [129] Y. Yarif et al., "Intra-nuclear cascade models at low energy?", in *Proc. 2007 Int. Conf. Nucl. Data Sci. Technol.*, Nice, EDP Sciences, 2008.
- [130] T. Sato, L. Sihver, H. Iwase, H. Nakashima and K. Niita, "Simulations of an accelerator-based shielding experiment using the particle and heavy-ion transport code system PHITS", *Adv. Space Res.*, vol. 35, no. 2, pp. 208-213, 2003.
- [131] W. A. Coleman and T. W. Armstrong, "Nucleon-meson transport code NMTC", Oak Ridge National Laboratory Report ORNL-4606, 1970.
- [132] A. Fassò, A. Ferrari, J. Ranft and P. R. Sala, "FLUKA: performances and applications in the intermediate energy range", in: *Proc. AEN/NEA Specialists' Meeting on Shielding Aspects of Accelerators, Targets and Irradiation Facilities*, OECD Documents, Paris, pp. 287-304, 1995.
- [133] M. Blann, "Hybrid model for pre-equilibrium decay in nuclear reactions", *Phys. Rev. Lett.*, vol. 27, pp. 337-340, 1971.
- [134] A. Ferrari and P. R. Sala, "The Physics of High Energy Reactions", in *Proc. Workshop Nucl. React. Data Nucl. React. Phys., Design and Safety*, A. Gandini, G. Re'ò eds., Trieste, 1998.
- [135] F. Ballarini et al., "Nuclear Models in FLUKA: Present Capabilities, Open Problems, and Future Improvements", *AIP Conf. Proc.*, vol. 769, pp. 1197-1202, 2005.
- [136] T. Koi et al., "Validation of Hadronic Models in Geant4", *AIP Conf. Proc.*, vol. 896, pp. 21-30, 2007.
- [137] J. P. Wellisch, "Geant4 hadronic physics status and validation for large HEP detectors", in *Proc. of CHEP 2003 Int. Conf. on Comp. in High En. and Nucl. Phys.*, La Jolla, 2003.
- [138] IEEE Computer Society, "IEEE Standard for Software Verification and Validation", IEEE Std 1012-2004, Jun. 2005.
- [139] United States Department of Energy, "Advanced Simulation and Computing Program Plan", Sandia National Laboratories Report SAND 2004-4607PP, Albuquerque, 2004.
- [140] J. Bisplinghoff, "Configuration mixing in preequilibrium reactions: A new look at the hybrid-exciton controversy", *Phys. Rev. C*, vol. 33, pp. 1569-1580, 1986.
- [141] A. J. Koning and M. C. Duijvestijn, "A global pre-equilibrium analysis from 7 to 200 MeV based on the optical model potential", *Nucl. Phys. A*, vol. 744, pp. 15-76, 2004.
- [142] G. Barrand, P. Binko, M. Donszelmann, A. Johnson, and A. Pfeiffer, "Abstract interfaces for data analysis: component architecture for data analysis tools", in *Proc. of CHEP Int. Conf. on Computing in High Energy and Nucl. Phys.*, pp. 215-218, Science Press, Beijing, 2001.
- [143] A. Pfeiffer, "iAIDA - an implementation of AIDA in C++". Online. Available: <http://iaida.dynalias.net/>
- [144] G. A. P. Cirrone et al., "A Goodness-of-Fit Statistical Toolkit", *IEEE Trans. Nucl. Sci.*, vol. 51, no. 5, pp. 2056-2063, 2004.
- [145] B. Mascialino, A. Pfeiffer, M. G. Pia, A. Ribon, and P. Viarengo, "New developments of the Goodness-of-Fit Statistical Toolkit", *IEEE Trans. Nucl. Sci.*, vol. 53, no. 6, pp. 3834-3841, 2006.
- [146] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain goodness of fit criteria based on stochastic processes", *Anls. Ma. St.*, vol. 23, pp. 193-212, 1952.
- [147] T. W. Anderson and D. A. Darling, "A test of goodness of fit", *JASA*, vol. 49, pp. 765-769, 1954.
- [148] H. Cramér, "On the composition of elementary errors. Second paper: statistical applications", *Skand. Aktuarietidskr.*, vol. 11, pp. 13-74, pp. 141-180, 1928.
- [149] R. von Mises, "Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik", Leipzig: F. Duticke, 1931.
- [150] M. Fisž, "On a result by M. Rosenblatt concerning the von Mises-Smirnov test", *Anls. Ma. St.*, vol. 31, pp. 427-429, 1960.
- [151] A. N. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione", *Gior. Ist. Ital. Attuari*, vol. 4, pp. 83-91, 1933.
- [152] N. V. Smirnov, "On the estimation of the discrepancy between empirical curves of distributions for two independent samples", *Bull. Math. Univ. Moscou*, 1939.
- [153] J. Kempe and A. Brahme, "Energy-range relation and mean energy variation in therapeutic particle beams", *Med. Phys.*, vol. 35, no.1, pp. 159-170, 2008.
- [154] P. M. DeLuca Jr., A. Wambersie, and G. Whitmore, "Prescribing, Recording, and Reporting Proton-Beam Therapy: Dosimetry", *J. ICRU*, vol. 7, no. 2, pp. 49-81, 2007.
- [155] P. M. DeLuca, Jr., A. Wambersie, and G. Whitmore, "Prescribing, Recording, and Reporting Proton-Beam Therapy: Beam Delivery and Properties", *J. ICRU*, vol. 7, no. 2, pp. 29-48, 2007.
- [156] L. J. Verhey et al., "The determination of absorbed dose in a proton beam for purposes of charged-particle radiation therapy", *Radiat. Res.*, vol. 79, pp. 34-54, 1979.
- [157] H. Paul, "The mean ionization potential of water, and its connection to the range of energetic carbon ions in water", *Nucl. Instrum. Meth. B*, vol. 255, pp. 435-437, 2007.
- [158] R. Schulte et al., "Conceptual Design of a Proton Computed Tomography System for Applications in Proton Radiation Therapy", *IEEE Trans. Nucl. Sci.*, vol. 51, no. 3, pp. 866-872, 2004.
- [159] H. Bichsel and T. Hiraoka, "Energy loss of 70 MeV protons in elements", *Nucl. Instrum. Meth. B*, vol. 66, pp. 345-351, 1992.
- [160] M. Dingfelder, "Cross section calculations in condensed media: charged particles in liquid water", *Radiat. Prot. Dosim.*, vol. 99, no. 1-4, pp. 23 - 28, 2002.
- [161] P. Andreo, "On the clinical spatial resolution achievable with protons and heavier charged particle radiotherapy beams", *Phys. Med. Biol.*, vol. 54, pp. N205-N215, 2009.
- [162] J. Soltani-Nabipour, D. Sardari, G. Cata-Danil, "Sensitivity of the Bragg peak curve to the average ionization potential of the stopping medium", *Rom. J. Phys.*, vol. 54, no. 3-4, pp. 321-330, 2009.
- [163] F. Sommerer, K. Parodi, A. Ferrari, K. Poljanc, W. Enghardt, and H. Aiginger, "Investigating the accuracy of the FLUKA code for transport of therapeutic ion beams in matter", *Phys. Med. Biol.*, vol. 51, pp. 4385-4398, 2006.
- [164] M. J. van Goethem et al., "Geant4 simulations of proton beam transport through a carbon or beryllium degrader and following a beam line", *Phys. Med. Biol.*, vol. 54, pp. 5831-5846, 2009.
- [165] Y. Kumazaki, T. Akagi, T. Yanou, D. Suga, Y. Hishikawa, and T. Teshima, "Determination of the mean excitation energy of water from proton beam ranges", *Radiat. Meas.*, vol. 42, pp. 1683-1691, 2007.
- [166] H. Paganetti, H. Jiang, K. Parodi, R. Slopsema, and M. Engelsman, "Clinical implementation of full Monte Carlo dose calculation in proton beam therapy", *Phys. Med. Biol.*, vol. 53, pp. 4825-4853, 2008.
- [167] I. Pshenichnov, A. Botvina, I. Mishustin, and W. Greiner, "Nuclear fragmentation reactions in extended media studied with Geant4 toolkit", submitted to *Nucl. Instrum. Meth. B*, Online. Available: <http://arxiv.org/pdf/0911.2017>.
- [168] J. F. Ziegler, J. P. Biersack, and M. D. Ziegler, "SRIM The Stopping and Range of Ions in Matter", Ed.: SRIM Co., Chester, MD, 2008.
- [169] H. Paul and A. Schinner, "Judging the reliability of stopping power tables and programs for protons and alpha particles using statistical methods", *Nucl. Instrum. Meth. B*, vol. 227, 461-470, 2005.
- [170] H. Paul, "A comparison of recent stopping power tables for light and medium-heavy ions with experimental data, and applications to radiotherapy dosimetry", *Nucl. Instrum. Meth. B*, vol. 247, 167-172, 2006.
- [171] A. Wald and J. Wolfowitz, "An exact test for randomness in the non-parametric case, based on serial correlation", *Ann. Math. Stat.*, vol. 14, pp. 378-388, 1943.
- [172] B. Gottschalk, R. Platais and H. Paganetti, "Nuclear interactions of for 160 MeV protons stopping in copper: a test of Monte Carlo nuclear models", *Med. Phys.*, vol. 26, no. 12, pp. 2597-2601, Dec. 1999.
- [173] C. Zacharatou Jarlskog and H. Paganetti, "Risk of Developing Second Cancer From Neutron Dose in Proton Therapy as Function of Field Characteristics, Organ, and Patient Age", *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 72, no. 1, pp. 228-235, 2008.
- [174] S. Guatelli et al., "Experience with Software Process in Physics Projects", in *Conf. Rec. 2004 IEEE Nucl. Sci. Symp.*, N40-8.
- [175] I. Jacobson, J. Booch, and J. Rumbaugh, "The Unified Software Development Process", Ed: Addison-Wesley, Boston, 1999.
- [176] M. A. Stephens, "EDF statistics for goodness-of-fit and some comparisons", *J. Amer. Statist. Assoc.*, vol. 69, pp. 730-737, 1974.
- [177] K. Pearson, "Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia", *Phil. Trans. Royal Soc. London, Series A*, vol. 187, pp. 253-318, 1896.

- [178] V. N. Ivanchenko, O. Kadri, M. Maire, and L. Urban, “Geant4 models for simulation of multiple scattering”, *J. Phys. Conf. Ser.*, vol. 219, pp. 032045, 2010.
- [179] G. J. Lockwood et al., “Calorimetric Measurement of Electron energy deposition in Extended Media - Theory vs Experiment”, Sandia National Laboratories, Report S AND79-0414, Albuquerque, 1980.
- [180] G. J. Lockwood et al., “Electron Energy and Charge Albedos - Calorimetric Measurement vs Monte Carlo Theory”, Sandia National Laboratories Report SAND80-1968, Albuquerque, 1981.
- [181] J. A. Halbleib and T. A. Mehlhorn, “ITS : The Integrated TIGER Series of Coupled Electron/Photon Monte Carlo Transport Codes”, Sandia National Laboratories Report No. SAND84-0573, Albuquerque, November 1984.